

PR 20-JAN-1999; 99US-0234395.
 PR 08-MAR-1999; 99US-0264516.
 PA (ADHE-) ADHEREX TECHNOLOGIES INC.
 XX
 XX
 PI Blaschuk OW, Gour BJ, Byers S;
 XX
 DR WPI; 2000-038791/03.
 XX
 PT New cadherin modulating agents, used for modulating nonclassical
 PT cadherin-mediated functions for treating e.g. cancers, obesity,
 PT rheumatoid arthritis, multiple sclerosis, diabetes or a neurological
 PT disease
 XX
 PS Claim 24; Page 163; 252pp; English.
 XX

CC The present invention describes cadherin modulating agents (MA) comprising peptides which comprise a nonclassical cadherin cell adhesion recognition (CAR) sequence. The MAs can be used for modulating nonclassical cadherin-mediated functions. They can be used for e.g. inhibiting adhesion of nonclassical-cadherin expressing cells in a mammal, enhancing delivery of a drug through the skin of a mammal, enhancing delivery of a drug to a tumour in a mammal, treating cancer in a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-expressing cell, preventing or treating obesity in a mammal, stimulating blood vessel regression in a mammal, enhancing drug delivery to the central nervous system, treating a demyelinating neurological disease, increasing vasopermeability in a mammal, enhancing adhesion of nonclassical cadherin-expressing cells, inhibiting synaptic stability in a mammal, or preventing pregnancy in a mammal. They can also be used for e.g. enhancing or directing neurite outgrowth, facilitating wound healing or reducing scar tissue, or enhancing adhesion of foreign tissue in a mammal. They can also be used for treating e.g. psoriasis, arthritis, age-related macular degeneration, multiple sclerosis and diabetes. The products can also be used for detection and diagnosis and in bioreactors. AAY60592 to AAY64572 represent specifically claimed peptides, and AAY64573 to AAY64643 and AA23183 to AA23186 represent sequences used in the exemplification of the present invention.

SO Sequence 5 AA;

Query Match 100.0%; Score 33; DB 21; Length 5;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 |||||
 1 CDAEC 5

RESULT 2

ID ABBG27768 standard; Protein; 467 AA.

XX ABBG27768;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #27759.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS91955.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 20; SEQ ID NO 58127; 103pp; English.
 XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (I) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABB00010-ABB30377 represent novel human CC diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pot_sequences.

SO Sequence 467 AA;

Query Match 100.0%; Score 33; DB 22; Length 467;
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 |||||
 11 CDAEC 15

RESULT 3

ID ABB59847 standard; Protein; 735 AA.

XX ABB59847;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 6333.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI, 2001-656860/75.
 XX N-PSDB; ABL03950.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 6333; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
 CC sequences (ABBS7737-ABBS72072).
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 735 AA;
 XX
 QY
 1 CDAEC 5
 |||||
 Db 658 CDAEC 662
 XX
 RESULT 4
 ABB71226
 ID ABB71226 standard; Protein; 744 AA.
 XX
 AC ABB71226;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 40470.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 PA Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI, 2001-656860/75.
 XX
 DR N-PSDB; ABL15329.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 40470; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
 CC sequences (ABBS7737-ABBS72072).
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 744 AA;
 XX
 QY
 1 CDAEC 5
 |||||
 Db 716 CDAEC 720
 XX
 RESULT 5
 ABB60498
 ID ABB60498 standard; Protein; 1679 AA.
 XX
 AC ABB60498;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 8286.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 PA Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI, 2001-656860/75.
 XX
 DR N-PSDB; ABL04601.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 8286; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
 CC sequences (ABBS7737-ABBS72072).
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 1679 AA;
 XX

Query Match 100.0%; Score 33; DB 22; Length 1679;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
DB 967 CDAC 971

RESULT 6
ID ABB60502 standard; Protein, 1679 AA.
XX ABB60502;
AC ABB60502;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 8298.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL04605.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 8298; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1679 AA;

Query Match 100.0%; Score 33; DB 22; Length 1679;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
DB 967 CDAC 971

RESULT 7
ID AAY62777 standard; Peptide; 5 AA.
XX

XX AAY62777;
AC
XX 02-MAR-2000 (first entry)
XX
XX PB-cadherin cell adhesion recognition cyclic peptide SEQ ID NO:2333.
XX
XX
XX Modulation; nonclassical cadherin mediated cell adhesion; CAR;
XX inhibition; cadherin extracellular domain; cell adhesion recognition;
XX OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8;
XX cadherin-12; cadherin-14; cadherin-15; T-cadherin; PB-cadherin;
XX cadherin related neuronal receptor; L1-cadherin; protocadherin;
XX desmoglein; desmocollin; calcium binding; cancer; tumor; obesity;
XX rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;
XX neurological disease; cyclic.
XX
XX Synthetic.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 1..5
XX
XX WO9957149-A2.
XX 11-NOV-1999.
XX 05-MAY-1999; 99WO-CA00363.
XX 05-MAY-1998; 98US-0073040.
XX 06-NOV-1998; 98US-0187859.
XX 20-JAN-1999; 99US-0234395.
XX 08-MAR-1999; 99US-0264516.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
XX Blaschuk OW, Gour BJ, Byers S;
XX WPI; 2000-038791/03.
XX
XX New cadherin modulating agents, used for modulating nonclassical
XX PT cadherin-mediated functions for treating e.g. cancers, obesity,
XX PT rheumatoid arthritis, multiple sclerosis, diabetes or a neurological
XX disease -
XX
XX Claim 72; Page 193; 252pp; English.

CC The present invention describes cadherin modulating agents (MA)
CC comprising peptides which comprise a nonclassical cadherin cell adhesion
CC recognition (CAR) sequence. The MAs can be used for modulating
CC nonclassical cadherin-mediated functions. They can be used for e.g.
CC inhibiting adhesion of nonclassical-cadherin expressing cells in a
CC mammal, enhancing delivery of a drug through the skin of a mammal,
CC enhancing delivery of a drug to a tumor in a mammal, treating cancer in
CC a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting
CC angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-
CC expressing cell, preventing or treating obesity in a mammal, stimulating
CC blood vessel regression in a mammal, enhancing drug delivery to the
CC central nervous system, treating a demyelinating neurological disease,
CC increasing vasopermeability in a mammal, enhancing adhesion of
CC nonclassical cadherin-expressing cells, inhibiting synaptic stability in
CC a mammal, or preventing pregnancy in a mammal. They can also be used for
CC e.g. enhancing or directing neurite outgrowth, facilitating wound
CC healing or reducing scar tissue, or enhancing adhesion of foreign tissue
CC in a mammal. They can also be used for treating e.g. psoriasis,
CC arthritis, age-related macular degeneration, multiple sclerosis and
CC diabetes. The products can also be used for detection and diagnosis and
CC in bioreactors. AAY60592 to AAY64572 represent specifically claimed
CC peptides, and AAY64573 to AAY64643 and AA233183 to AA233186 represent
CC sequences used in the exemplification of the present invention.
XX
XX Sequence 5 AA;

Query Match 90.9%; Score 30; DB 21; Length 5;

Best Local Similarity 80.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DY 1 CDAEC 5
1 CDAEC 5
DB 1 CDAEC 5

RESULT 8

AAV64273

ID AAV64273 standard; Peptide; 5 AA.

XX AAV64273;

XX 02-MAR-2000 (first entry)

DE Cadherin-related neuronal receptor CAR cyclic peptide SEQ ID NO:3587.

XX Modulation; nonclassical cadherin mediated cell adhesion; CAR;
XX Inhibition; cadherin extracellular domain; cell adhesion recognition;
XX OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8;
XX cadherin-12; cadherin-14; cadherin-15; T-cadherin; PB-cadherin;
XX cadherin related neuronal receptor; LI-cadherin; protocadherin;
XX desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;
XX rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;
XX neurological disease; cyclic.

XX Synthetic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Disulfide-bond 1..5

XX WO957149-A2.

XX 11-NOV-1999.

XX 05-MAY-1999; 99WO-CA00363.

XX 05-MAY-1998; 98US-0073040.

XX 06-NOV-1998; 98US-0187859.

XX 20-JAN-1999; 99US-0234395.

XX 08-MAR-1999; 99US-0264516.

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuk OW, Gour BJ, Byers S;

XX WPI; 2000-038791/03.

XX New cadherin modulating agents, used for modulating nonclassical
XX cadherin-mediated functions for treating e.g. cancers, obesity,
XX rheumatoid arthritis, multiple sclerosis, diabetes or a neurological
XX disease

XX Claim 102; Page 217; 252pp; English.

XX The present invention describes cadherin modulating agents (MA)
XX comprising peptides which comprise a nonclassical cadherin cell adhesion
XX recognition (CAR) sequence. The MAs can be used for modulating
XX nonclassical cadherin-mediated functions. They can be used for e.g.
XX inhibiting adhesion of nonclassical-cadherin expressing cells in a
XX mammal, enhancing delivery of a drug through the skin of a mammal,
XX enhancing delivery of a drug to a tumour in a mammal, treating cancer in
XX a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting
XX angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-
XX expressing cell, preventing or treating obesity in a mammal, stimulating
XX blood vessel regression in a mammal, enhancing drug delivery to the
XX central nervous system, treating a demyelinating neurological disease,
XX increasing vasopermeability in a mammal, enhancing adhesion of
XX nonclassical cadherin-expressing cells, inhibiting synaptic stability in
XX a mammal, or preventing pregnancy in a mammal. They can also be used for
XX e.g. enhancing or directing neurite outgrowth, facilitating wound

CC healing or reducing scar tissue, or enhancing adhesion of foreign tissue
CC in a mammal. They can also be used for treating e.g. psoriasis,
CC arthritis, age-related macular degeneration, multiple sclerosis and
CC diabetes. The products can also be used for detection and diagnosis and
CC in bioreactors. AAV60592 to AAV64572 represent specifically claimed
CC peptides, and AAV64573 to AAV64643 and AA233183 represent
XX sequences used in the exemplification of the present invention.

SQ Sequence 5 AA;

Query Match Score 30; DB 21; Length 5;

Best Local Similarity 80.0%; Pred. No. 7.8e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;DY 1 CDAEC 5
1 CDAEC 5
DB 1 CDAEC 5

RESULT 9

ID AAR1372 standard; Protein; 40 AA.

XX AAR1372;

XX 08-MAY-1991 (first entry)

DE Exon II encoded by genomic meg-CSF clone.

XX Megakaryocyte colony stimulating factor; platelet deficiency;
XX bleeding disorder.

XX Homo sapiens.

XX WO9102001-A.

XX 21-FEB-1991.

XX 07-AUG-1990; 90WO-US04421.

XX 29-JUN-1990; 90US-0546114.

XX 08-AUG-1989; 89US-0390901.

XX 28-DEC-1989; 89US-0457196.

XX (GENE-) GENETICS INST INC.

XX Gesner TG, Clark SC, Turner K, Hewick RM;

XX WPI; 1991-073490/10.

XX N-PSDB; AAQ10580.

XX New mega:karyocyte colony stimulating factor protein - regulates
XX human haematopoiesis by stimulating growth and development of
XX mega:karyocyte(s) in treatment of e.g. plastic anaemia

XX Claim 3; Page 85; 204pp; English.

XX The clone was isolated from a human placenta lambda phage DNA
XX library. The sequence can be inserted into expression vectors for
XX the prodn. of recombinant meg-CSF. The protein is used to treat
XX bleeding disorders and platelet deficiencies.
XX See also AAR10870, AAR10871 and AAR10872.

SQ Sequence 40 AA;

Query Match Score 30; DB 12; Length 40;

Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;DY 1 CDAEC 5
1 CDAEC 5
DB 20 CDAEC 24

RESULT 10
AAU41409
ID AAU41409 standard; Protein; 82 AA.
AC AAU41409;
XX
XX
DT 13-FEB-2002 (first entry)
XX
XX
DE Propionibacterium acnes immunogenic protein #2305.
XX
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopthc; neuroprotectant.
XX
XX
OS Propionibacterium acnes.
XX
XX
XX MO200181581-A2.
XX
XX
XX 01-NOV-2001.
XX
XX
XX 20-APR-2001; 2001MO-US12865.
XX
XX
XX 21-APR-2000; 2000US-199047P.
XX
XX 02-JUN-2000; 2000US-208841P.
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX
XX (CORI-) CORIAX CORP.
XX
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'italouneuve J, Zhang Y, Jen S, Carter D;
XX
XX
XX WPI: 2001-616774/71.
XX
XX
XX N-PSDB; AAS59515.
XX
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX
XX Example 1; SEQ ID No 2604; 10699p; English.
XX
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 82 AA;
XX
XX
XX Query Match 90.9%; Score 30; DB 22; Length 82;
XX Best Local Similarity 80.0%; Pred. No. 3.1e+02;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 1 CDAEC 5
XX
XX 16 CDADC 20

RESULT 11
AAR26050
ID AAR26050 standard; Protein; 111 AA.
XX
XX
XX AAR26050;
XX
XX
DT 02-FEB-1993 (first entry)
XX
XX
DE MSF-K130.
XX
XX
XX Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
XX stability; proteolytic cleavage; adhesion; alternative splicing;
XX MSF-precursor.
XX
XX
XX Synthetic.
XX
XX
XX
XX Key Location/Qualifiers
XX
XX Region 1..4
XX
XX /label= Enterokinase_site
XX
XX Region 7..48
XX
XX /label= Exon_II
XX
XX Region 48..88
XX
XX /label= Exon_III
XX
XX Region 88..111
XX
XX /label= Exon_IV
XX
XX
XX MO9213075-A.
XX
XX
XX 06-AUG-1992.
XX
XX
XX 17-JAN-1992; 92MO-US00433.
XX
XX
XX 18-JAN-1991; 91US-0643502.
XX
XX 10-SEP-1991; 91US-0757022.
XX
XX
XX (GENMY) GENETICS INST INC.
XX
XX
XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
XX
XX
XX WPI: 1992-284660/34.
XX
XX
XX N-PSDB; AAQ27224.
XX
XX
XX New human mega-karyocyte stimulating factors - for treating
XX immune deficiencies, cancer, exposure to radiation or drugs,
XX bacterial and viral infections, etc.
XX
XX
XX Disclosure; Fig 3; 87p; English.
XX
XX
XX The sequence covered by exons II, III and IV encodes megakaryocyte
XX stimulating factor (MSF). This sequence was used to produce MSF as
XX a fusion protein with thiodoxin. This sequence was derived from a
XX MSF-precursor (see also AAR26050).
XX
XX The precursor cDNA sequence contains sequences derived from human
XX megakaryocyte colony stimulating factor (meg-CSF). Exon I contains
XX the initiating methionine, and encodes a classical mammalian protein
XX secretion signal sequence. The sequence encoding the original meg-CSF
XX includes exons II-IV and is thought to terminate in the region between
XX amino acid residues 114 - 147. The primary transcript of this gene
XX may be cleaved in different ways to yield a family of mRNA's each
XX encoding a different MSF protein. Exons V and VI are thought to be
XX related to the activity of the factor and are also implicated in the
XX stability, folding and processing of the molecule. These exons are
XX also thought to play a role in the observed synergy of MSF with other
XX cytokines. Exons V - XII are believed to be implicated in the
XX processing or folding of the appropriate structure of the resulting
XX factor, ie. one or more of these exons may contain sequences which
XX direct proteolytic cleavage, adhesion, organisation of the cellular
XX matrix or extracellular matrix processing. Both naturally occurring
XX and non-naturally occurring MSF's may be characterised by various
XX combinations of alternatively spliced exons from this sequence, with
XX the exons spliced together in differing orders to form different
XX members of the MSF family.

XX Sequence 111 AA;
SQ

Query Match 90.9%; Score 30; DB 13; Length 111;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5
|||:
Db 67 CDAQC 71

RESULT 12

AAB94916
ID AAB94916 standard; Protein; 151 AA.

XX AAB94916;
AC

XX 26-JUN-2001 (first entry)
DT

XX Human protein sequence SEQ ID NO:16355.
XX

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX

XX Homo sapiens.
XX

XX EPI074617-A2.
XX

XX 07-FEB-2001.
XX

XX 28-JUL-2000; 2000EP-0116126.
XX

XX 29-JUL-1999; 99JP-0248036.
XX

XX 27-AUG-1999; 99JP-0300253.
XX

XX 11-JAN-2000; 2000JP-0118776.
XX

XX 02-MAY-2000; 2000JP-0183767.
XX

XX 09-JUN-2000; 2000JP-0241899.
XX

XX (HELI-) HELIX RES INST.
XX

XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

XX WPI; 2001-318749/34.
XX

XX Claim 8; SEQ ID 16355; 2537pp + CD ROM; English.
XX

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises at least 15 nucleotides, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.
XX

XX Sequence 151 AA;
SQ

Query Match 90.9%; Score 30; DB 22; Length 151;
Best Local Similarity 80.0%; Pred. No. 5.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5
|||:
Db 122 CDADC 126

RESULT 13

AAR10872
ID AAR10872 standard; Protein; 182 AA.

XX AAR10872;
AC

XX 08-MAY-1991 (first entry)
DT

XX Protein encoded by meg-CSF cDNA clone.
XX

XX Megakaryocyte colony stimulating factor; platelet deficiency;
XX bleeding disorder.
XX

XX Homo sapiens.
XX

XX WO9102001-A.
XX

XX 21-FEB-1991.
XX

XX 07-AUG-1990; 90WO-US04421.
XX

XX 29-JUN-1990; 90US-0546114.
XX

XX 08-AUG-1989; 89US-0390901.
XX

XX 28-DEC-1989; 89US-0457196.
XX

XX (GENE-) GENETICS INST INC.
XX

XX Gesner TG, Clark SC, Turner K, Hewick RM;
XX

XX WPI; 1991-073490/10.
XX

XX N-PSDB; AAQ10582.
XX

XX Claim 3; Page 85; 204pp; English.
XX

XX The clone was isolated from a human placenta lambda phage DNA
XX library. The sequence can be inserted into expression vectors for
XX the prodn. of recombinant meg-CSF. The protein is used to treat
XX bleeding disorders and platelet deficiencies.
XX See also AAR11372, AAR10870 and AAR10871.
XX

XX Sequence 182 AA;
SQ

Query Match 90.9%; Score 30; DB 12; Length 182;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5
|||:
Db 93 CDAQC 97

RESULT 14

AAU11261
ID AAU11261 standard; Protein; 292 AA.

XX AAU11261;
AC

XX 12-MAR-2002 (first entry)
 XX Human HAPO polypeptide.
 DE Human; HAPO, stimulating factor; haemopoietic stem ancestral cell;
 KW vascular endothelial cell; haemopoietic disease; vascular disease.
 XX Homo sapiens.
 OS
 XX CN1312294-A.
 PN
 XX 12-SEP-2001.
 PD
 XX 28-FEB-2001; 2001CN-0109083.
 PF
 XX 28-FEB-2001; 2001CN-0109083.
 PR
 XX (HEMA-) INST HEMATOLOGY CHINESE ACAD MEDICAL SCI.
 XX Han Z, Liu Y, Cai Y;
 XX WPI, 2002-018506/03.
 DR N-PSDB; AAS16924.
 DR
 XX Human blood and blood vessel cytopoiesis hormone and its preparation -
 PT
 XX Claim 1; Page 9 (disclosure); 20pp; Chinese.
 PS
 XX The invention relates to a stimulating factor, HAPO, with the effect of
 CC promoting growth of haemopoietic stem ancestral cells and vascular
 CC endothelial cells. HAPO is characterised using several pairs of primers
 CC specific for HAPO DNA to make a series of gene segments by PCR
 CC amplification. An antibody specific to the HAPO polypeptide can be used
 CC in the treatment and prevention of haemopoietic and vascular diseases.
 CC This sequence represents the HAPO polypeptide of the invention.
 CC
 SQ Sequence 292 AA;
 QY
 DB 1 CDAEC 5
 46 CDAQC 50
 LT 15
 AAB97171 standard; Protein; 341 AA.
 ID AAB97171;
 AC AAB97171;
 XX
 DT 13-AUG-2001 (first entry)
 XX
 DE Rainbow trout galectin.
 XX
 KW Rainbow trout; galectin; cytosolic; anti-inflammatory; immunity;
 KW fish disease; cell adhesion inhibition; inflammation; cancer;
 KW T cell movement; apoptosis.
 XX
 OS Oncochrychus mykiss.
 XX
 PN JP2001069976-A.
 XX
 PD 21-MAR-2001.
 PT
 XX 01-SEP-1999; 99JP-0247204.
 PF
 XX 01-SEP-1999; 99JP-0247204.
 PR
 XX (NORQ) NORINSUISANSHO YOSHOKU KENKYU.
 PA

XX WPI; 2001-321173/34.
 DR N-PSDB; AAB24641.
 DR
 XX Novel recombinant Rainbow trout galectin protein and gene encoding the
 PT protein, useful for studying fish immunity mechanism, and diagnosis of
 PT fish diseases -
 XX
 PS Claim 1; Page 10-11; 17pp; Japanese.
 XX
 CC The invention relates to a recombinant protein having a 341 amino acid
 CC sequence fully defined in the specification, or its mutant in which at
 CC least one amino acid is deleted, substituted or added, but which retains
 CC the galectin activity. The protein is useful for studying mechanisms of
 CC fish immunity and for diagnosing fish diseases. The galectin proteins are
 CC involved in inhibition of cell adhesion, inflammation, metastasis of
 CC tumour cells, T cell movement, and apoptosis. The present sequence
 CC is the protein of the invention.
 CC
 SQ Sequence 341 AA;
 QY
 DB 1 CDAEC 5
 298 CDAQC 302
 LT 16
 AAB83409 standard; protein; 402 AA.
 ID AAB83409
 AC AAB83409;
 XX
 DT 22-MAY-1996 (first entry)
 XX
 DE Humicola insolens endoglucanase variant Q399N..
 DE
 KW Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KW detergent; fabric softener; defibrillation; deparking; bio-polishing;
 KW de-inking; paper pulp drainage; animal feed; fabric washing.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers.
 FT Misc-difference 399
 FT Misc-difference /note="Q399N"
 PN W09524471-A1.
 PN
 PD 14-SEP-1995.
 PD
 XX 08-MAR-1995; 95WO-DK00108.
 PF
 XX 30-MAR-1994; 94DK-0000365.
 PR 08-MAR-1994; 94DK-0000270.
 PR
 XX (NOVO) NOVO-NORDISK AS.
 PA
 PI Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
 XX
 DR WPI; 1995-366105/47.
 DR
 PT Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.
 XX
 PS Claim 16; -; 76pp; English.
 XX
 CC AAB83405-R83419 represent mutations of the endoglucanase of Humicola
 CC insolens (the wild type sequence is represented by AAB83398).
 CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for

CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and
 CC bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibration or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The
 CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5

Db 172 CDAQC 176

AT 17

ID AAR83410 standard; protein; 402 AA.

AC AAR83410;

DT 22-MAY-1996 (first entry)

DE Humicola insolens endoglucanase variant E202A.

KW Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibration; debarking; bio-polishing;
 KM de-inking; paper pulp drainage; animal feed; fabric washing.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 202 /note= "E202A"

PN W09524471-A1.

PD 14-SEP-1995.

PF 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

PR 08-MAR-1994; 94DK-0000270.

(NOVO) NOVO-NORDISK AS.

PI Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;

XX WPI; 1995-366105/47.

DR Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.

XX Claim 16; -; 76pp; English.

XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola

CC insolens (the wild type sequence is represented by AAR83398).

CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for

CC soil removal in detergents, or in fabric softeners. They can also be

CC used for colour clarification of textile fabrics, for preventing

CC backstaining in fabric washing, for textile soil removal and

CC bio-polishing, for de-inking used paper, for paper pulp drainage

CC improvement and defibration or debarking. The proteins can also be used

CC in the baking and wine and juice industry or for animal feeds. The

CC mutated sequences have improved fibre wall penetrating ability, and can

CC provide more efficient degradation of cellulosic materials while not

CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;

Best Local Similarity 80.0%; Pred. No. 1.1e+03; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5

Db 172 CDAQC 176

RESULT 18

ID AAR83411 standard; protein; 402 AA.

AC AAR83411;

DT 22-MAY-1996 (first entry)

DE Humicola insolens endoglucanase variant S37W + P39W.

KW Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibration; debarking; bio-polishing;
 KM de-inking; paper pulp drainage; animal feed; fabric washing.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 37 /note= "S37W"

FT Misc-difference 39 /note= "P39W"

PN W09524471-A1.

PD 14-SEP-1995.

PF 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

PR 08-MAR-1994; 94DK-0000270.

(NOVO) NOVO-NORDISK AS.

PI Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;

XX WPI; 1995-366105/47.

DR Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.

XX Claim 16; -; 76pp; English.

XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola

CC insolens (the wild type sequence is represented by AAR83398).

CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for

CC soil removal in detergents, or in fabric softeners. They can also be

CC used for colour clarification of textile fabrics, for preventing

CC backstaining in fabric washing, for textile soil removal and

CC bio-polishing, for de-inking used paper, for paper pulp drainage

CC improvement and defibration or debarking. The proteins can also be used

CC in the baking and wine and juice industry or for animal feeds. The

CC mutated sequences have improved fibre wall penetrating ability, and can

CC provide more efficient degradation of cellulosic materials while not

CC reducing the tensile strength of fabrics.

QY 1 CDAEC 5
 DB 172 CDAQC 176

RESULT 19
 AAR83412

ID AAR83412 standard; protein; 402 AA.

XX AAR83412;

XX 22-MAY-1996 (first entry)

XX Humicola insolens endoglucanase variant M142E.

XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibrillation; debarking; bio-polishing;
 KW de-inking; paper pulp drainage; animal feed; fabric washing.

XX Synthetic.

PH Key Location/Qualifiers
 FT Misc-difference 142
 FT /note= "M142E"

XX MO9524471-A1.

XX 14-SEP-1995.

XX 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

XX 08-MAR-1994; 94DK-0000270.

XX (NOVO) NOVO-NORDISK AS.

XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;

XX WPI; 1995-366105/47.

XX Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.

XX Claim 17; -: 76pp; English.

XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
 CC insolens (the wild type sequence is represented by AAR83398).

XX Endoglucanase is a family 7 cellulase. These cellulases can be used for
 CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and

XX bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibrillation or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The

CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 172 CDAQC 176

RESULT 20

XX AAR83413 standard; protein; 402 AA.

XX

AC AAR83413;
 XX 22-MAY-1996 (first entry)

XX Humicola insolens endoglucanase variant K217A.

XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibrillation; debarking; bio-polishing;
 KW de-inking; paper pulp drainage; animal feed; fabric washing.

XX Synthetic.

PH Key Location/Qualifiers
 FT Misc-difference 217
 FT /note= "K217A"

XX MO9524471-A1.

XX 14-SEP-1995.

XX 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

XX 08-MAR-1994; 94DK-0000270.

XX (NOVO) NOVO-NORDISK AS.

XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;

XX WPI; 1995-366105/47.

XX Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.

XX Claim 17; -: 76pp; English.

XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
 CC insolens (the wild type sequence is represented by AAR83398).

XX Endoglucanase is a family 7 cellulase. These cellulases can be used for
 CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and

XX bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibrillation or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The

CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 172 CDAQC 176

RESULT 21

XX AAR83414 standard; protein; 402 AA.

XX AAR83414;

XX 22-MAY-1996 (first entry)

XX Humicola insolens endoglucanase variant K217A + K218T.

XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibrillation; debarking; bio-polishing;
 KW de-inking; paper pulp drainage; animal feed; fabric washing.

```

XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 217 /note= "K217A"
XX FT Misc-difference 218 /note= "K218T"
XX
XX PN WO9524471-A1.
XX
XX PD 14-SEP-1995.
XX
XX PF 08-MAR-1995; 95WO-DK00108.
XX
XX PR 30-MAR-1994; 94DK-0000365.
XX PR 08-MAR-1994; 94DK-0000270.
XX
XX PA (NOVO ) NOVO-NORDISK AS.
XX
XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
XX WPI, 1995-366105/47.
XX
XX PT Modified Family 7 cellulase enzymes - used for soil removal in
XX detergent compns. or in fabric softener compns.
XX
XX PS Claim 17; -: 76pp; English.
XX
XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
XX insolens (the wild type sequence is represented by AAR83398).
XX Endoglucanase is a Family 7 cellulase. These cellulases can be used for
XX soil removal in detergents, or in fabric softeners. They can also be
XX used for colour clarification of textile fabrics, for preventing
XX backstaining in fabric washing, for textile soil removal and
XX bio-polishing, for de-inking used paper. The proteins can also be used
XX improvement and defibration or debarking. The proteins can also be used
XX in the baking and wine and juice industry or for animal feeds. The
XX mutated sequences have improved fibre wall penetrating ability, and can
XX provide more efficient degradation of cellulosic materials while not
XX reducing the tensile strength of fabrics.
XX
XX SQ Sequence 402 AA;
XX
XX Query Match 90.9%; Score 30; DB 16; Length 402;
XX Best Local Similarity 80.0%; Pred. No. 1.1e+03;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 CDAEC 5
XX |||:|
XX 172 CDAQC 176
XX
XX RESULT 22
XX AAR83415
XX ID AAR83415 standard; protein; 402 AA.
XX
XX AC AAR83415;
XX
XX DT 22-MAY-1996 (first entry)
XX
XX DE Humicola insolens endoglucanase variant R245G.
XX
XX KM Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
XX detergent; fabric softener; defibration; debarking; bio-polishing;
XX de-inking; paper pulp drainage; animal feed; fabric washing.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 245 /note= "R245G"
XX
XX

```

```

PN WO9524471-A1.
XX
XX PD 14-SEP-1995.
XX
XX PF 08-MAR-1995; 95WO-DK00108.
XX
XX PR 30-MAR-1994; 94DK-0000365.
XX PR 08-MAR-1994; 94DK-0000270.
XX
XX PA (NOVO ) NOVO-NORDISK AS.
XX
XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
XX WPI, 1995-366105/47.
XX
XX PT Modified Family 7 cellulase enzymes - used for soil removal in
XX detergent compns. or in fabric softener compns.
XX
XX PS Claim 17; -: 76pp; English.
XX
XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
XX insolens (the wild type sequence is represented by AAR83398).
XX Endoglucanase is a Family 7 cellulase. These cellulases can be used for
XX soil removal in detergents, or in fabric softeners. They can also be
XX used for colour clarification of textile fabrics, for preventing
XX backstaining in fabric washing, for textile soil removal and
XX bio-polishing, for de-inking used paper, for paper pulp drainage
XX improvement and defibration or debarking. The proteins can also be used
XX in the baking and wine and juice industry or for animal feeds. The
XX mutated sequences have improved fibre wall penetrating ability, and can
XX provide more efficient degradation of cellulosic materials while not
XX reducing the tensile strength of fabrics.
XX
XX SQ Sequence 402 AA;
XX
XX Query Match 90.9%; Score 30; DB 16; Length 402;
XX Best Local Similarity 80.0%; Pred. No. 1.1e+03;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 CDAEC 5
XX |||:|
XX 172 CDAQC 176
XX
XX RESULT 23
XX AAR83416
XX ID AAR83416 standard; protein; 402 AA.
XX
XX AC AAR83416;
XX
XX DT 22-MAY-1996 (first entry)
XX
XX DE Humicola insolens endoglucanase variant I310D.
XX
XX KM Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
XX detergent; fabric softener; defibration; debarking; bio-polishing;
XX de-inking; paper pulp drainage; animal feed; fabric washing.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 310 /note= "I310D"
XX
XX PN WO9524471-A1.
XX
XX PD 14-SEP-1995.
XX
XX PF 08-MAR-1995; 95WO-DK00108.
XX
XX PR 30-MAR-1994; 94DK-0000365.
XX PR 08-MAR-1994; 94DK-0000270.
XX
XX

```


CC improvement and defibrillation or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The
 CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 172 CDAQC 176

RESULT 26

AA83398 ID AAR83398 standard; protein; 402 AA.

XX AAR83398;

22-MAY-1996 (first entry)

XX Humicola insolens endoglucanase.

XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibrillation; debarking; bio-polishing;
 KM de-inking; paper pulp drainage; animal feed; fabric washing.

XX Humicola insolens.

XX MO9524471-A1.

XX 14-SEP-1995.

XX 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

XX 08-MAR-1994; 94DK-0000270.

XX (NOVO) NOVO-NORDISK AS.

XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
 XX WPI; 1995-366105/47.

XX Modified Family 7 cellulase enzymes - used for soil removal in
 XX detergent compns. or in fabric softener compns.

XX Claim 5; Fig 3; 76pp; English.

XX This sequence represents endoglucanase, a family 7 cellulase isolated
 CC from Humicola insolens. The sequence was mutated to give the sequences
 CC represented by AAR83405-R83419. The mutated cellulases can be used for
 CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and
 CC bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibrillation or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The
 CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5

DB 172 CDAQC 176

RESULT 27

AA83405 ID AAR83405 standard; protein; 402 AA.

XX AAR83405;

22-MAY-1996 (first entry)

XX Humicola insolens endoglucanase variant N89Q.

XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibrillation; debarking; bio-polishing;
 KM de-inking; paper pulp drainage; animal feed; fabric washing.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 89 /note= "N89Q"

XX MO9524471-A1.

XX 14-SEP-1995.

XX 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

XX 08-MAR-1994; 94DK-0000270.

XX (NOVO) NOVO-NORDISK AS.

XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
 XX WPI; 1995-366105/47.

XX Modified Family 7 cellulase enzymes - used for soil removal in
 XX detergent compns. or in fabric softener compns.

XX Claim 16; -; 76pp; English.

XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
 CC insolens (the wild type sequence is represented by AAR83398).
 CC Endoglucanase is a family 7 cellulase. These cellulases can be used for
 CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and
 CC bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibrillation or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The
 CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

XX Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 172 CDAQC 176

RESULT 28

AA83406 ID AAR83406 standard; protein; 402 AA.

XX AAR83406;

```

XX 22-MAY-1996 (first entry)
XX Humicola insolens endoglucanase variant N89Q + N247Q.
DE Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
XX detergent; fabric softener; defibrillation; debarking; bio-polishing;
KM de-inking; paper pulp drainage; animal feed; fabric washing.
XX Synthetic.
XX Key Location/Qualifiers
FH MISC-difference 89 /note= "N89Q"
FT MISC-difference 247 /note= "N247Q"
FT MISC-difference 247 /note= "N247Q"
XX W09524471-A1.
XX 14-SEP-1995.
XX 08-MAR-1995; 95WO-DK00108.
XX 30-MAR-1994; 94DK-0000365.
XX 08-MAR-1994; 94DK-0000270.
XX (NOVO ) NOVO-NORDISK AS.
XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
XX WPI; 1995-366105/47.
XX Modified Family 7 cellulase enzymes - used for soil removal in
PT detergent compns. or in fabric softener compns.
XX Claim 16; -: 76pp; English.
XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
CC insolens (the wild type sequence is represented by AAR83398).
CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for
CC soil removal in detergents, or in fabric softeners. They can also be
CC used for colour clarification of textile fabrics, for preventing
CC bio-polishing, for de-inking used paper, for paper pulp drainage
CC improvement and defibrillation or debarking. The proteins can also be used
CC in the baking and wine and juice industry or for animal feeds. The
CC mutated sequences have improved fibre wall penetrating ability, and can
CC provide more efficient degradation of cellulosic materials while not
CC reducing the tensile strength of fabrics.
XX SQ Sequence 402 AA;
XX
XX Query Match 90.9%; Score 30; DB 16; Length 402;
XX Best Local Similarity 80.0%; Pred. No. 1.1e+03;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 CDAEC 5
XX 172 CDAQC 176
XX
XX Db
XX
XX RESULT 29
XX AAR83407
XX ID AAR83407 standard; protein; 402 AA.
XX AAR83407;
XX 22-MAY-1996 (first entry)
XX Humicola insolens endoglucanase variant H123N.
XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
XX detergent; fabric softener; defibrillation; debarking; bio-polishing;
XX de-inking; paper pulp drainage; animal feed; fabric washing.
XX W09524471-A1.
XX 14-SEP-1995.
XX 08-MAR-1995; 95WO-DK00108.
XX 30-MAR-1994; 94DK-0000365.
XX 08-MAR-1994; 94DK-0000270.
XX (NOVO ) NOVO-NORDISK AS.
XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
XX WPI; 1995-366105/47.
XX Modified Family 7 cellulase enzymes - used for soil removal in
PT detergent compns. or in fabric softener compns.
XX Claim 16; -: 76pp; English.
XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
CC insolens (the wild type sequence is represented by AAR83398).
CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for
CC soil removal in detergents, or in fabric softeners. They can also be
CC used for colour clarification of textile fabrics, for preventing
CC bio-polishing, for de-inking used paper, for paper pulp drainage
CC improvement and defibrillation or debarking. The proteins can also be used
CC in the baking and wine and juice industry or for animal feeds. The
CC mutated sequences have improved fibre wall penetrating ability, and can
CC provide more efficient degradation of cellulosic materials while not
CC reducing the tensile strength of fabrics.
XX SQ Sequence 402 AA;
XX
XX Query Match 90.9%; Score 30; DB 16; Length 402;
XX Best Local Similarity 80.0%; Pred. No. 1.1e+03;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 CDAEC 5
XX 172 CDAQC 176
XX
XX Db
XX
XX RESULT 30
XX AAR83408
XX ID AAR83408 standard; protein; 402 AA.
XX AAR83408;
XX 22-MAY-1996 (first entry)
XX Humicola insolens endoglucanase variant T385N.
XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
XX detergent; fabric softener; defibrillation; debarking; bio-polishing;
XX de-inking; paper pulp drainage; animal feed; fabric washing.
XX Synthetic.
XX Key Location/Qualifiers
FH MISC-difference 385 /note= "T385N"
FT MISC-difference 385 /note= "T385N"
XX W09524471-A1.

```

```

KM de-inking; paper pulp drainage; animal feed; fabric washing.
XX Synthetic.
XX Key Location/Qualifiers
FH MISC-difference 123 /note= "H123N"
FT MISC-difference 123 /note= "H123N"
XX W09524471-A1.
XX 14-SEP-1995.
XX 08-MAR-1995; 95WO-DK00108.
XX 30-MAR-1994; 94DK-0000365.
XX 08-MAR-1994; 94DK-0000270.
XX (NOVO ) NOVO-NORDISK AS.
XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
XX WPI; 1995-366105/47.
XX Modified Family 7 cellulase enzymes - used for soil removal in
PT detergent compns. or in fabric softener compns.
XX Claim 16; -: 76pp; English.
XX AAR83405-R83419 represent mutations of the endoglucanase of Humicola
CC insolens (the wild type sequence is represented by AAR83398).
CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for
CC soil removal in detergents, or in fabric softeners. They can also be
CC used for colour clarification of textile fabrics, for preventing
CC bio-polishing, for de-inking used paper, for paper pulp drainage
CC improvement and defibrillation or debarking. The proteins can also be used
CC in the baking and wine and juice industry or for animal feeds. The
CC mutated sequences have improved fibre wall penetrating ability, and can
CC provide more efficient degradation of cellulosic materials while not
CC reducing the tensile strength of fabrics.
XX SQ Sequence 402 AA;
XX
XX Query Match 90.9%; Score 30; DB 16; Length 402;
XX Best Local Similarity 80.0%; Pred. No. 1.1e+03;
XX Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 CDAEC 5
XX 172 CDAQC 176
XX
XX Db
XX
XX RESULT 30
XX AAR83408
XX ID AAR83408 standard; protein; 402 AA.
XX AAR83408;
XX 22-MAY-1996 (first entry)
XX Humicola insolens endoglucanase variant T385N.
XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
XX detergent; fabric softener; defibrillation; debarking; bio-polishing;
XX de-inking; paper pulp drainage; animal feed; fabric washing.
XX Synthetic.
XX Key Location/Qualifiers
FH MISC-difference 385 /note= "T385N"
FT MISC-difference 385 /note= "T385N"
XX W09524471-A1.

```

XX 14-SEP-1995.
 PD 08-MAR-1995; 95WO-DK00108.
 XX
 PF 30-MAR-1994; 94DK-0000365.
 PR 08-MAR-1994; 94DK-0000270.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
 XX WPI, 1995-366105/47.
 DR
 XX Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.
 XX
 PS Claim 16; -; 76pp; English.
 XX
 CC AAR83405-R83419 represent mutations of the endoglucanase of Humicola
 CC insoliens (the wild type sequence is represented by AAR83398).
 CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for
 CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and
 CC bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibration or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The
 CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.
 CC
 SQ Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5
 |||:
 Db 172 CDAQC 176

RESULT 31

AAR83419
 ID AAR83419 standard; protein; 402 AA.

XX
 AC AAR83419;

XX
 22-MAY-1996 (first entry)

Humicola insolens endoglucanase variant M198L.

XX Endoglucanase; humicola insolens; family 7 cellulase; soil removal;
 KM detergent; fabric softener; defibration; debarking; bio-polishing;
 KM de-inking; paper pulp drainage; animal feed; fabric washing.

XX Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 198
 FT /note= "M198L"

XX W09524471-A1.

XX 14-SEP-1995.

XX 08-MAR-1995; 95WO-DK00108.

XX 30-MAR-1994; 94DK-0000365.

XX 08-MAR-1994; 94DK-0000270.

XX (NOVO) NOVO-NORDISK AS.

XX Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;
 PI WPI, 1995-366105/47.
 XX
 DR Modified Family 7 cellulase enzymes - used for soil removal in
 PT detergent compans. or in fabric softener compans.
 XX
 PS Claim 17; -; 76pp; English.
 XX

CC AAR83405-R83419 represent mutations of the endoglucanase of Humicola
 CC insoliens (the wild type sequence is represented by AAR83398).
 CC Endoglucanase is a Family 7 cellulase. These cellulases can be used for
 CC soil removal in detergents, or in fabric softeners. They can also be
 CC used for colour clarification of textile fabrics, for preventing
 CC backstaining in fabric washing, for textile soil removal and
 CC bio-polishing, for de-inking used paper, for paper pulp drainage
 CC improvement and defibration or debarking. The proteins can also be used
 CC in the baking and wine and juice industry or for animal feeds. The
 CC mutated sequences have improved fibre wall penetrating ability, and can
 CC provide more efficient degradation of cellulosic materials while not
 CC reducing the tensile strength of fabrics.

SQ Sequence 402 AA;

Query Match 90.9%; Score 30; DB 16; Length 402;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDAEC 5
 |||:
 Db 172 CDAQC 176

RESULT 32

AAW01501
 ID AAW01501 standard; protein; 409 AA.

XX
 AC AAW01501;

XX 26-FEB-1997 (first entry)

XX 50 kD endoglucanase, EG I-F.

XX Detergent composition; cellulase; retaining-type activity; catalytic
 KM activity; cellobiose; particulate soil removal; colour clarification;
 KM cleaning; cellulose-containing fabric; cellulohydrolase; endoglucanase;

XX Fusarium oxysporum, DSM 2672.

XX W09502675-A1.

XX 26-JAN-1995.

XX 07-JUL-1994; 94WO-DK00280.

XX 11-OCT-1993; 93DK-0001135.

XX 12-JUL-1993; 93EP-0870131.

XX (NOVO) NOVO-NORDISK AS.

XX (PROC) PROCTER & GAMBLE CO.

XX Convents AC, Jeffreys B, Schuelein M, Tikhomirov DF;

XX WPI, 1995-067325/09.

XX Detergent compans. contg. two cellulase components - the first
 PT removing soil particles and the second capable of colour
 PT clarification, useful in laundry compans.

XX Claim 19; Page 63-65; 83pp; English.
 CC Detergent compositions comprising: 1) a first cellulase component having

CC retaining-type activity, pref. having a catalytic activity on cellobiose
 CC at pH 8.5 corresponding to Kcat of at least 0.01 s⁻¹, and capable of
 CC particulate soil removal; and 2) a second cellulase component having
 CC multiple domains comprising at least one non-catalytic domain attached to
 CC a catalytic domain, pref. having a catalytic activity on Red AVICel 7.5
 CC per 1 mg of cellulase protein higher than 10⁻⁴ IU and being capable of
 CC colour clarification, where at least one of the cellulase components is a
 CC single (recombinant) component, are useful for cleaning and colour
 CC clarification of cellulose-containing fabrics. The first cellulase
 CC component can be an endoglucanase which is immunoreactive with an
 CC antibody raised against a highly purified ~50 kD endoglucanase
 CC derived from Fusarium oxysporum, DSM 2672, and is esp. the present
 CC sequence, designated EG I-F.

XX Sequence 409 AA;

Query Match 90.9%; Score 30; DB 16; Length 409;

Best Local Similarity 80.0%; Pred. No. 1.1e+03; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CDAEC 5

170 CDAQC 174

RESULT 33

AAW17927 standard; Protein; 409 AA.

AAW17927;

08-AUG-1997 (first entry)

Abbrading endoglucanase.

Denim; desizing; stone-washing; endoglucanase; cellulase.

Fusarium oxysporum strain DSM 2672.

MO9718286-A1.

22-MAY-1997.

15-NOV-1996; 96WO-DK00469.

15-NOV-1995; 95DK-0001278.

(NOVO) NOVO-NORDISK AS.

Lund H;

WPI; 1997-289265/26.

Combined desizing and stone-washing of dyed denim - by treatment
 PT with an amyolytic enzyme, an abrading endoglucanase and a
 XX streak-reducing endoglucanase

Claim 10; Page 18-20; 31pp; English.

XX A novel one-step process for combining desizing and 'stone-washing'
 CC of dyed denim involves treatment with an amyolytic enzyme in
 CC combination with a first abrading monocomponent endoglucanase, such
 CC as the EG V type cellulases of Humicola insolens DSM 1800 (AAW17926)
 CC or Fusarium oxysporum DSM 2672 (AAW17927), and a second streak-
 CC reducing monocomponent endoglucanase, such as the endoglucanase of
 CC H. insolens DSM 1800 (AAW17928). Improved colour uniformity and
 CC distribution and stone-wash quality are obtd. The denim may
 CC additionally be treated with a thermostable lipolytic enzyme.

Sequence 409 AA;

Query Match 90.9%; Score 30; DB 18; Length 409;

Best Local Similarity 80.0%; Pred. No. 1.1e+03;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CDAEC 5

170 CDAQC 174

RESULT 34

AAW38850 standard; Protein; 410 AA.

AAW38850;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 1995.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

peripheral nervous system; neuropathy; central nervous system; CNS;

Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

leukaemia.

Homo sapiens.

26-DEC-2000; 2000WO-US344263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

Zhao QH, Zhou P, Goodrich R, Drmanac RT;

WPI; 2001-442253/47.

N-PSDB; AAI58006.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Example 3; SEQ ID NO 1995; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and

the encoded polypeptides (AAW38850-AAW42213) with nootropic,

immunosuppressant and cytostatic activity. The polynucleotides are useful

in gene therapy. A composition containing a polypeptide or polynucleotide

of the invention may be used to treat diseases of the peripheral nervous

system, such as peripheral nervous injuries, peripheral neuropathy and

localised neuropathies and central nervous system diseases, such as

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

utilisation of the activities such as: Immune system suppression,

Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, cancer diagnosis and therapy, drug screening,

assays for receptor activity, arthritis and inflammation, leukaemias and

C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed

specification.

Sequence 410 AA;

Query Match. 90.9%; Score 30; DB 22; Length 410;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 15 CDSEC 19

RESULT 35

AA83399 standard; protein; 411 AA.

AA83399;
 22-MAY-1996 (first entry)
 Fusarium oxysporum endoglucanase.
 Endoglucanase; fusarium oxysporum; family 7 cellulase; soil removal;
 detergent; fabric softener; defibrillation; debarking; bio-polishing;
 de-inking; paper pulp drainage; animal feed; fabric washing.

Fusarium oxysporum.

MO9524471-A1.

14-SEP-1995.

08-MAR-1995; 95WO-DK00108.

30-MAR-1994; 94DK-0000365.

08-MAR-1994; 94DK-0000270.

(NOVO) NOVO-NORDISK AS.

Hansen SA, Nielsen JB, Rosholm P, Schuelein M, Von Der Osten C;

WPI; 1995-366105/47.

Modified Family 7 cellulase enzymes - used for soil removal in
 detergent compns. or in fabric softener compns.

Claim 6; Fig 4; 76pp; English.

This sequence represents endoglucanase, a family 7 cellulase isolated
 from Fusarium oxysporum. Mutated versions of this sequence are
 referred to in claims 16, 17 and 18. These mutations include N89Q,
 N247Q, H123N, T385N, Q399N, E202A, S37W, P39W, M142E, K217N, K218T,
 R245G, I310D, E150Q, E334K and M198L. However the numbering of these
 mutants appears to be based on the numbering of the Humicola insolens
 endoglucanase (see AA83398), and is such that the mutants of this
 sequence can not be created. This cellulase (and mutated versions of it)
 can be used for soil removal in detergents, or in fabric softeners. They
 can also be used for colour clarification of textile fabrics, for
 preventing backstaining in fabric washing, for textile soil removal and
 bio-polishing, for de-inking used paper, for paper pulp drainage
 improvement and delamination or debarking. The proteins can also be used
 in the baking and wine and juice industry or for animal feeds. The
 mutated sequences have improved fibre wall penetrating ability, and can
 provide more efficient degradation of cellulosic materials while not
 reducing the tensile strength of fabrics.

Sequence 411 AA;

Query Match 90.9%; Score 30; DB 16; Length 411;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 172 CDAQC 176

RESULT 36

AAE12787 standard; Protein; 414 AA.

AAE12787;
 15-JAN-2002 (first entry)

Talaromyces emersonii beta-glucanase CEB protein.

Beta-glucanase; CEB; antipilemic; fungicide; hyperlipemia; animal feed;
 fabric; edible foodstuff; textile; brewing; distilling; bioremediation;
 dental hygiene; leather-treatment; paper manufacture; plant extract;
 detergent treatment; baking.

Talaromyces emersonii.

MO200170998-A1.

27-SEP-2001.

20-MAR-2001; 2001MO-EP03174.

20-MAR-2000; 2000EP-0302263.

(STAM) DSM NV.

Van Den Hombergh JPTW, Van Der Laan J, Daraan JG, Herweijer MA;

Teufel DP;

WPI; 2001-648392/74.

N-PSDB; AAD20929.

New polypeptide for treating hyperlipemia and/or high serum
 cholesterol and triglyceride levels, comprises the beta-glucanase
 protein obtainable from Talaromyces fungus -

Claim 3; Page 69-70; 76pp; English.

The invention relates to a beta-glucanase polypeptide obtainable from a
 fungus of the genus Talaromyces, e.g. T. emersonii, having endoglucanase
 activity (EC 3.2.1.4). Beta-glucanase polypeptide is useful for the
 manufacture of a medicament for treating hyperlipemia, high serum
 cholesterol and triglyceride levels. Beta-glucanase polypeptide are
 useful for treating fungal or plant materials (plant pulp, plant
 extracts), edible foodstuffs or ingredients, or fabrics, textiles or
 clothes containing plant materials. Beta-glucanase polypeptide is
 useful for reducing the viscosity of a plant material, for cleaving
 beta-D-glucan polymers in the plant material, and for processing plant
 pulp, juice or extract, by incubating the pulp, juice or extract.
 Beta-glucanase polypeptide is applicable in brewing, distilling,
 bioremediation, dental hygiene, leather-treatment, paper manufacture,
 textile treatment or manufacture, baking or bread making, washing or
 detergent treatment, treating flower bulbs or in animal feed.
 Beta-glucanase polypeptide is also useful during the production of milk
 substituted from soybean. The present sequence is Talaromyces emersonii
 beta-glucanase CEB protein.

Sequence 414 AA;

Query Match 90.9%; Score 30; DB 22; Length 414;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 191 CDAQC 195

RESULT 37

AA67390

PT Combined desizing and stone-washing of dyed denim - by treatment
 PT with an amylolytic enzyme, an abrading endoglucanase and a
 PT streak-reducing endoglucanase
 XX
 PS Claim 9, Page 16-18; 31pp; English.
 CC A novel one-step process for combining desizing and 'stone-washing'
 CC of dyed denim involves treatment with an amylolytic enzyme in
 CC combination with a first abrading monocomponent endoglucanase, such
 CC as the BG V type cellulases of Humicola insolens DSM 1800 (AAW17926)
 CC or Fusarium oxysporum DSM 2672 (AAW17927), and a second streak-
 CC reducing monocomponent endoglucanase, such as the endoglucanase of
 CC H. insolens DSM 1800 (AAW17928). Improved colour uniformity and
 CC distribution and stone-wash quality are obtd. The denim may
 CC additionally be treated with a thermostable lipolytic enzyme.
 XX
 SQ Sequence 415 AA;
 Query Match 90.9%; Score 30; DB 18; Length 415;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 172 CDAQC 176
 1 CDABC 5
 |||:
 |||:
 |||:
 RESULT 40
 AAY57056
 ID AAY57056 standard; protein; 415 AA.
 XX
 AC AAY57056;
 XX
 DT 21-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of cellulase EGI without a cellulose binding domain.
 XX
 KM Cellulase; cellulolytic enzyme; EGI; cellulase binding domain; CBD;
 KM laundry detergent; fabric care composition; strain removal; fabric colour;
 KM fabric whiteness; anti-bobbling; depilling; fabric softness; clean.
 XX
 OS Accomycota.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FH Misc-difference 61..92.
 FT /label= Unknown
 FT /note= "These residues represent a line of missing
 FT text in the sequence listing in the
 FT specification. They are included to maintain the
 FT residue numbering given in the specification for
 FT this protein sequence"
 FT Misc-difference 301..316
 FT /label= Unknown
 FT /note= "These residues represent a line of missing
 FT text in the sequence listing in the
 FT specification. They are included to maintain the
 FT residue numbering given in the specification for
 FT this protein sequence"
 XX
 PN WO957256-A1.
 XX
 PD 11-NOV-1999.
 XX
 XX 01-MAY-1998; 98WO-US08903.
 XX
 PR 01-MAY-1998; 98WO-US08903.
 XX
 PA (PROC) PROCTER & GAMBLE CO.
 XX
 PT Busch A, Bettiol JP, Smets J, Boyer SL;
 XX
 DR WPI; 2000-052967/04.

XX
 PT Modified enzyme for laundry detergents and/or fabric care compositions
 PT for e.g. cleaning, strain removal, whiteness maintenance, fabric
 PT softness, colour appearance and fabric anti-wear properties -
 XX
 PS Disclosure; Page 77-79; 87pp; English.
 CC This is the amino acid sequence of the cellulolytic enzyme EGI, without a
 CC cellulose binding domain (CBD). The enzyme is used in a laundry detergent
 CC and/or fabric care composition, which is made from this catalytically
 CC active EGI amino acid sequence linked to an amino acid sequence of a CBD.
 CC The modified EGI enzyme of the invention is derived from Humicola
 CC insolens or Trichoderma reesei, has an approximate molecular weight of
 CC 50kD, an isoelectric point of 5.5 and contains the EGI 415 amino acid
 CC sequence linked to a CBD. The enzyme is useful in laundry detergent
 CC and/or fabric care compositions which can be used for cleaning, fabric
 CC stain removal, fabric whiteness maintenance; anti-bobbling, depilling,
 CC fabric softness, colour appearance and fabric anti-wear properties and
 CC benefits A detergent made with the enzyme has improved cleaning benefits,
 CC without any undesirable impact on the fabric, even at high enzymatic
 CC concentrations.
 XX
 SQ Sequence 415 AA;
 Query Match 90.9%; Score 30; DB 21; Length 415;
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 172 CDAQC 176
 1 CDABC 5
 |||:
 |||:
 |||:
 OY
 DB 172 CDAQC 176

Search completed: July 11, 2003, 15:28:24
 Job time : 37 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 15:27:46 ; Search time 27 Seconds
(without alignments)
5.449 Million cell updates/sec

Title: US-10-006-869-910

Perfect score: 33

Sequence: 1 CDAEC 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues 262574

Total number of hits satisfying chosen parameters: 262574
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_Aa:*
1: /cgn2_6/prodata/1/aa/5A COMB pep.*
2: /cgn2_6/prodata/1/aa/5B COMB pep.*
3: /cgn2_6/prodata/1/aa/6A COMB pep.*
4: /cgn2_6/prodata/1/aa/6B COMB pep.*
5: /cgn2_6/prodata/1/aa/PCFUS COMB pep.*
6: /cgn2_6/prodata/1/aa/backfile1 pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	5	US-09-187-859-910	Sequence 910, App
2	30	90.9	5	US-09-187-859-910	Sequence 233, App
3	30	90.9	5	US-09-187-859-910	Sequence 3587, App
4	30	90.9	40	US-07-757-022B-8	Sequence 8, App1
5	30	90.9	97	US-07-757-022B-36	Sequence 36, App1
6	30	90.9	106	US-07-757-022B-56	Sequence 56, App1
7	30	90.9	111	US-07-757-022B-28	Sequence 28, App1
8	30	90.9	115	US-07-757-022B-108	Sequence 108, App
9	30	90.9	115	US-07-757-022B-122	Sequence 122, App
10	30	90.9	116	US-07-757-022B-98	Sequence 98, App1
11	30	90.9	116	US-07-757-022B-110	Sequence 110, App
12	30	90.9	116	US-07-757-022B-120	Sequence 120, App
13	30	90.9	116	US-07-757-022B-138	Sequence 138, App
14	30	90.9	130	US-07-757-022B-78	Sequence 78, App1
15	30	90.9	130	US-07-757-022B-86	Sequence 86, App1
16	30	90.9	131	US-07-757-022B-82	Sequence 82, App1
17	30	90.9	132	US-07-757-022B-140	Sequence 140, App
18	30	90.9	138	US-07-757-022B-34	Sequence 34, App1
19	30	90.9	141	US-07-757-022B-80	Sequence 80, App1
20	30	90.9	147	US-07-757-022B-38	Sequence 38, App1
21	30	90.9	156	US-07-757-022B-106	Sequence 106, App
22	30	90.9	156	US-07-757-022B-118	Sequence 118, App
23	30	90.9	157	US-07-757-022B-114	Sequence 114, App
24	30	90.9	157	US-07-757-022B-102	Sequence 102, App
25	30	90.9	172	US-07-757-022B-88	Sequence 88, App1
26	30	90.9	188	US-07-757-022B-32	Sequence 32, App1
27	30	90.9	192	US-07-757-022B-90	Sequence 90, App1

28	30	90.9	204	US-07-757-022B-92	Sequence 92, App1
29	30	90.9	207	US-07-757-022B-116	Sequence 116, App
30	30	90.9	207	US-07-757-022B-136	Sequence 136, App
31	30	90.9	208	US-07-757-022B-132	Sequence 132, App
32	30	90.9	209	US-07-757-022B-94	Sequence 94, App1
33	30	90.9	220	US-07-757-022B-96	Sequence 96, App1
34	30	90.9	231	US-07-757-022B-30	Sequence 30, App1
35	30	90.9	237	US-07-757-022B-72	Sequence 72, App1
36	30	90.9	296	US-07-757-022B-70	Sequence 70, App1
37	30	90.9	372	US-07-757-022B-64	Sequence 64, App1
38	30	90.9	377	US-08-169-948B-14	Sequence 14, App1
39	30	90.9	377	US-08-448-873-14	Sequence 14, App1
40	30	90.9	377	US-08-382-452D-14	Sequence 14, App1
41	30	90.9	402	US-08-709-979A-3	Sequence 3, App1
42	30	90.9	402	US-08-709-974A-1	Sequence 1, App1
43	30	90.9	402	US-08-709-974A-5	Sequence 5, App1
44	30	90.9	409	US-08-709-974A-3	Sequence 3, App1
45	30	90.9	409	US-09-069-632-2	Sequence 2, App1

ALIGNMENTS

```
RESULT 1
US-09-187-859-910
; Sequence 910, Application US/09187859A
; Patent No. 6358920
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C1
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 910
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: cadherin-5 cell adhesion recognition sequence
US-09-187-859-910

Query Match      100.0%; Score 33; DB 4; Length 5;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 CDAEC 5
Db      1 CDAEC 5

RESULT 2
US-09-187-859-2333
; Sequence 2333, Application US/09187859A
; Patent No. 6358920
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C1
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2333
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

FEATURE:
OTHER INFORMATION: Representative cyclic modulating agent based on
OTHER INFORMATION: PB-cadherin cell adhesion recognition sequence
US-09-187-859-2333

Query Match 90.9%; Score 30; DB 4; Length 5;
Best Local Similarity 80.0%; Pred. No. 2e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 1 CDADC 5

RESULT 3

US-09-187-859-3587
Sequence 3587, Application US/09187859A
Patent No. 6358920

GENERAL INFORMATION:

APPLICANT: Blaschuk, Orest W.

APPLICANT: Gour, Barbara J.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
FILE REFERENCE: 100086.407C1

CURRENT APPLICATION NUMBER: US/09/187,859A

CURRENT FILING DATE: 1998-11-06

NUMBER OF SEQ ID NOS: 4052

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 3587

LENGTH: 5

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Representative cyclic modulating agent based on
OTHER INFORMATION: cadherin-related neuronal receptor cell adhesion
OTHER INFORMATION: recognition sequence

US-09-187-859-3587

Query Match 90.9%; Score 30; DB 4; Length 5;
Best Local Similarity 80.0%; Pred. No. 2e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 1 CDADC 5

RESULT 4

US-07-757-022B-8
Sequence 8, Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Gesner, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 40 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-8

Query Match 90.9%; Score 30; DB 4; Length 40;
Best Local Similarity 80.0%; Pred. No. 52;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 20 CDADC 24

RESULT 5

US-07-757-022B-36
Sequence 36, Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-36

Query Match 90.9%; Score 30; DB 4; Length 97;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 CDAC 5
45 CDAQ 49
DB

RESULT 6
US-07-757-022B-56
Sequence 56, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geeser, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-56

Query Match 90.9%; Score 30; DB 4; Length 106;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
DB 86 CDAQ 90

RESULT 7
US-07-757-022B-28
Sequence 28, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geeser, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 111 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-28

Query Match 90.9%; Score 30; DB 4; Length 111;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CDAC 5
Db 67 CDAQC 71

RESULT 8

US-07-757-022B-108
Sequence 108, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 108:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-108

Query Match 90.9%; Score 30; DB 4; Length 115;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDAC 5
Db 45 CDAQC 49

RESULT 9
US-07-757-022B-122
Sequence 122, Application US/07757022B

Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-122

Query Match 90.9%; Score 30; DB 4; Length 115;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDAC 5
Db 45 CDAQC 49

RESULT 10
US-07-757-022B-98
Sequence 98, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-98

Query Match 90.9%; Score 30; DB 4; Length 116;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDABC 5
DB 45 CDAQC 49

US-07-757-022B-110
Sequence 110, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-110

Query Match 90.9%; Score 30; DB 4; Length 116;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDABC 5
DB 45 CDAQC 49

RESULT 12
US-07-757-022B-120
Sequence 120, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
7-757-022B-120

Query Match 90.9%; Score 30; DB 4; Length 116;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 45 CDAQC 49

RESULT 13
US-07-757-022B-138
Sequence 138, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-138

Query Match 90.9%; Score 30; DB 4; Length 116;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 45 CDAQC 49

RESULT 14
US-07-757-022B-78
Sequence 78, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 130 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-78

Query Match 90.9%; Score 30; DB 4; Length 130;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 15

US-07-757-022B-86
; Sequence 86, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 130 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; US-07-757-022B-86

Query Match 90.9%; Score 30; DB 4; Length 130;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 16

US-07-757-022B-82
; Sequence 82, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; US-07-757-022B-82

Query Match 90.9%; Score 30; DB 4; Length 131;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 61 CDAQC 65

RESULT 17

US-07-757-022B-140
; Sequence 140, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 140:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-140

Query Match 90.9%; Score 30; DB 4; Length 132;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 CDAEC 5
61 CDAQC 65

RESULT 18
US-07-757-022B-34
Sequence 34, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-34

Query Match 90.9%; Score 30; DB 4; Length 138;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDAEC 5
Db 86 CDAQC 90

RESULT 19
US-07-757-022B-80
Sequence 80, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
07-757-022B-80

Query Match 90.9%; Score 30; DB 4; Length 141;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 20

US-07-757-022B-38
Sequence 38, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geeser, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 147 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-38

Query Match 90.9%; Score 30; DB 4; Length 147;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 45 CDAQC 49

RESULT 21

US-07-757-022B-106
Sequence 106, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geeser, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-106

Query Match 90.9%; Score 30; DB 4; Length 156;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:|
DB 86 CDAQC 90

RESULT 22
US-07-757-022B-118
Sequence 118, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-118

Query Match 90.9%; Score 30; DB 4; Length 156;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
|||:|

DB 86 CDAQC 90

RESULT 23
US-07-757-022B-102
Sequence 102, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 157 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-102

Query Match 90.9%; Score 30; DB 4; Length 157;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
|||:|
DB 86 CDAQC 90

RESULT 24
US-07-757-022B-114
Sequence 114, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 157 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-114

Query Match 90.9%; Score 30; DB 4; Length 157;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 25
US-07-757-022B-88
Sequence 88, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.

ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 172 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-88

Query Match 90.9%; Score 30; DB 4; Length 172;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 26
US-07-757-022B-32
Sequence 32, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-32

Query Match 90.9%; Score 30; DB 4; Length 188;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
|||:|
Db 86 CDAQC 90

RESULT 27
US-07-757-022B-90
Sequence 90, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
INFORMATION FOR SEQ ID NO: 92:

APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-90

Query Match 90.9%; Score 30; DB 4; Length 192;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
|||:|
Db 86 CDAQC 90

RESULT 28
US-07-757-022B-92
Sequence 92, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 92:

SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-92

Query Match 90.9%; Score 30; DB 4; Length 204;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 29
US-07-757-022B-116
Sequence 116, Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 116:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-116

Query Match 90.9%; Score 30; DB 4; Length 207;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 30
US-07-757-022B-136
Sequence 136, Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-136

Query Match 90.9%; Score 30; DB 4; Length 207;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 31
US-07-757-022B-132
Sequence 132, Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 208 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
07-757-022B-132

Query Match 90.9%; Score 30; DB 4; Length 208;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:|
Db 86 CDAQC 90

RESULT 32
US-07-757-022B-94
Sequence 94, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-94

Query Match 90.9%; Score 30; DB 4; Length 209;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:|
Db 86 CDAQC 90

RESULT 33
US-07-757-022B-96
Sequence 96, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-96

Query Match 90.9%; Score 30; DB 4; Length 220;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 34
US-07-757-022B-30
Sequence 30, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 231 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-30

Query Match 90.9%; Score 30; DB 4; Length 231;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 35
US-07-757-022B-72
Sequence 72, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-72

Query Match 90.9%; Score 30; DB 4; Length 237;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 86 CDAQC 90

ULT 36
US-07-757-022B-70

Sequence 70; Application US/07757022B
Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Gesner, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 70:

SEQUENCE CHARACTERISTICS:

LENGTH: 296 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

Query Match 90.9%; Score 30; DB 4; Length 296;
Best Local Similarity 80.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 86 CDAQC 90

RESULT 37

US-07-757-022B-64

Sequence 64; Application US/07757022B

Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Gesner, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-64

Query Match 90.9%; Score 30; DB 4; Length 372;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 45 CDAQC 49

RESULT 38

US-08-169-948B-14
Sequence 14, Application US/08169948B
Patent No. 5861271
GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen
APPLICANT: Collier, Katherine
APPLICANT: Larenaas, Edmund
TITLE OF INVENTION: No. 5861271el Cellulase Enzymes and Systems
TITLE OF INVENTION: For Their Expression
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,948B
FILING DATE: DEC 17 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
REFERENCE/DOCKET NUMBER: GC226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7536
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-948B-14

Query Match 90.9%; Score 30; DB 2; Length 377;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
171 CDAQC 175

RESULT 39
US-08-448-873-14
Sequence 14, Application US/08448873
Patent No. 5874276
GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen
APPLICANT: Collier, Katherine A.
APPLICANT: Larenaas, Edmund
TITLE OF INVENTION: No. 5874276el Cellulase Enzymes and Systems
TITLE OF INVENTION: For Their Expression
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,873
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/169,948
FILING DATE: 17-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher L.
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC226D14
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-448-873-14

Query Match 90.9%; Score 30; DB 2; Length 377;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 171 CDAQC 175

RESULT 40
US-08-382-452D-14
Sequence 14, Application US/08382452D
Patent No. 6268196
GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Ward, Michael
APPLICANT: Collier, Katherine D.
APPLICANT: Larenaas, Edmund A.
TITLE OF INVENTION: NOVEL CELLULOSE ENZYMES AND SYSTEMS
TITLE OF INVENTION: FOR THEIR EXPRESSION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/382,452D
FILING DATE: February 1, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 36,696
REFERENCE/DOCKET NUMBER: GC226-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-382-452D-14

Query Match 90.9%; Score 30; DB 4; Length 377;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 171 CDAQC 175

Search completed: July 11, 2003, 15:31:50
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 11, 2003, 15:29:51 / Search time 22 Seconds
(without alignments)
26.459 Million cell updates/sec

Title: US-10-006-869-910

Perfect score: 1 CDAEC 5

Sequence: 1 CDAEC 5

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PC7_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/PCUS_PUBCOMB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	33	100.0	5	9	US-10-006-869-910
2	30	90.9	5	9	US-10-006-869-2333
3	30	90.9	5	9	US-10-006-869-3587
4	30	90.9	40	12	US-10-124-557-8
5	30	90.9	57	12	US-10-124-557-36
6	30	90.9	106	12	US-10-124-557-56
7	30	90.9	111	12	US-10-124-557-28
8	30	90.9	115	12	US-10-124-557-108
9	30	90.9	115	12	US-10-124-557-122
10	30	90.9	116	12	US-10-124-557-98
11	30	90.9	116	12	US-10-124-557-110
12	30	90.9	116	12	US-10-124-557-120
13	30	90.9	130	12	US-10-124-557-138
14	30	90.9	130	12	US-10-124-557-78
15	30	90.9	130	12	US-10-124-557-86
16	30	90.9	131	12	US-10-124-557-82
17	30	90.9	132	12	US-10-124-557-140
18	30	90.9	138	12	US-10-124-557-34
19	30	90.9	141	12	US-10-124-557-80

20	30	90.9	147	12	US-10-124-557-38	Sequence 38, Appl
21	30	90.9	156	12	US-10-124-557-106	Sequence 106, Appl
22	30	90.9	156	12	US-10-124-557-118	Sequence 118, Appl
23	30	90.9	157	12	US-10-124-557-102	Sequence 102, Appl
24	30	90.9	157	12	US-10-124-557-114	Sequence 114, Appl
25	30	90.9	172	12	US-10-124-557-88	Sequence 88, Appl
26	30	90.9	188	12	US-10-124-557-32	Sequence 32, Appl
27	30	90.9	192	12	US-10-124-557-90	Sequence 90, Appl
28	30	90.9	204	12	US-10-124-557-92	Sequence 92, Appl
29	30	90.9	207	12	US-10-124-557-116	Sequence 116, Appl
30	30	90.9	207	12	US-10-124-557-136	Sequence 136, Appl
31	30	90.9	208	12	US-10-124-557-132	Sequence 132, Appl
32	30	90.9	209	12	US-10-124-557-94	Sequence 94, Appl
33	30	90.9	220	12	US-10-124-557-96	Sequence 96, Appl
34	30	90.9	231	12	US-10-124-557-30	Sequence 30, Appl
35	30	90.9	237	12	US-10-124-557-72	Sequence 72, Appl
36	30	90.9	296	12	US-10-124-557-70	Sequence 70, Appl
37	30	90.9	372	12	US-10-124-557-64	Sequence 64, Appl
38	30	90.9	377	9	US-09-916-494A-14	Sequence 14, Appl
39	30	90.9	422	12	US-10-124-557-68	Sequence 68, Appl
40	30	90.9	428	7	US-08-841-636A-33	Sequence 33, Appl
41	30	90.9	452	7	US-08-841-636A-35	Sequence 35, Appl
42	30	90.9	463	12	US-10-124-557-54	Sequence 54, Appl
43	30	90.9	1022	12	US-10-124-557-84	Sequence 84, Appl
44	30	90.9	1038	12	US-10-124-557-74	Sequence 74, Appl
45	30	90.9	1049	12	US-10-124-557-58	Sequence 58, Appl

ALIGNMENTS

RESULT 1
US-10-006-869-910
; Sequence 910, Application US/10006869
; Publication No. US20030082166A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; FILE REFERENCE: 100086.407C7
; CURRENT APPLICATION NUMBER: US/10/006,869
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 910
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: cadherin-5 cell adhesion recognition sequence
US-10-006-869-910

Query Match 100.0%; Score 33; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 1 CDAEC 5

RESULT 2
US-10-006-869-2333
; Sequence 2333, Application US/10006869
; Publication No. US20030082166A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL

FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-124-557-36

Query Match 90.9%; Score 30; DB 12; Length 97;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 45 CDAQC 49

RESULT 6
US-10-124-557-56
Sequence 56, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
City: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-10-124-557-56

Query Match 90.9%; Score 30; DB 12; Length 106;
Best Local Similarity 80.0%; Pred. No. 2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 86 CDAQC 90

RESULT 7
US-10-124-557-28
Sequence 28, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
City: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 111 amino acids
TYPE: amino acid

TOPOLOGY: linear.
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-124-557-28

Query Match 90.9%; Score 30; DB 12; Length 111;
Best local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 67 CDAQC 71

RESULT 8

US-10-124-557-108
Sequence 108, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-Jan-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-Jun-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-Dec-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-Aug-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:

LENGTH: 115 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 108:

US-10-124-557-108

QY 1 CDAEC 5

Db 45 CDAQC 49

RESULT 9

US-10-124-557-122
Sequence 122, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-Jan-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-Jun-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-Dec-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-Aug-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 122:

SEQUENCE CHARACTERISTICS:

LENGTH: 115 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 122:

US-10-124-557-122

QY 1 CDAEC 5

Db 45 CDAQC 49

RESULT 10

US-10-124-557-98
Sequence 98, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

QY 1 CDAEC 5

Db 45 CDAQC 49

Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 98:
US-10-124-557-98
Query Match 90.9%; Score 30; DB 12; Length 116;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 CDAEC 5
45 CDAQC 49
Db
RESULT 11
US-10-124-557-110
Sequence 110, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 110:
US-10-124-557-110
Query Match 90.9%; Score 30; DB 12; Length 116;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 CDAEC 5
45 CDAQC 49
Db
RESULT 12
US-10-124-557-120
Sequence 120, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 120:
US-10-124-557-120

Query Match 90.9%; Score 30; DB 12; Length 116;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 45 CDAQC 49

RESULT 13

US-10-124-557-138
Sequence 138, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 138:
US-10-124-557-138

Query Match 90.9%; Score 30; DB 12; Length 116;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 45 CDAQC 49

RESULT 14

US-10-124-557-78
Sequence 78, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 130 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-124-557-78

Query Match 90.9%; Score 30; DB 12; Length 130;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 15

US-10-124-557-86
; Sequence 86, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 130 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-10-124-557-86

Query Match 90.9%; Score 30; DB 12; Length 130;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 86 CDAQC 90

RESULT 16

US-10-124-557-82
; Sequence 82, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-10-124-557-82

Query Match 90.9%; Score 30; DB 12; Length 131;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 61 CDAQC 65

RESULT 17

US-10-124-557-140
; Sequence 140, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 140:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 140:
US-10-124-557-140
Query Match 90.9%; Score 30; DB 12; Length 132;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
61 CDAQC 65
RESULT 18
US-10-124-557-34
Sequence 34, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-124-557-34
Query Match 90.9%; Score 30; DB 12; Length 138;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
86 CDAQC 90
RESULT 19
US-10-124-557-80
Sequence 80, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196

FILED DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-10-124-557-80

Query Match 90.9%; Score 30; DB 12; Length 141;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDABC 5
Db 86 CDAQC 90

RESULT 20
US-10-124-557-38
Sequence 38, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:
LENGTH: 147 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-10-124-557-38

Query Match 90.9%; Score 30; DB 12; Length 147;
Best Local Similarity 80.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDABC 5
Db 45 CDAQC 49

RESULT 21
US-10-124-557-106
Sequence 106, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-10-124-557-106

Query Match 90.9%; Score 30; DB 12; Length 156;
Best Local Similarity 80.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 86 CDAQC 90

RESULT 22

US-10-124-557-118
Sequence 118, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT:

Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESSES:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:

LENGTH: 156 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 118:

US-10-124-557-118

Query Match 90.9%; Score 30; DB 12; Length 156;
Best Local Similarity 80.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 86 CDAQC 90

RESULT 23

US-10-124-557-102

Sequence 102, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT:

Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESSES:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 102:

SEQUENCE CHARACTERISTICS:

LENGTH: 157 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 102:

US-10-124-557-102

Query Match 90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 80.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 86 CDAQC 90

RESULT 24

US-10-124-557-114

Sequence 114, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT:

Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESSES:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 157 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 114:
US-10-124-557-114
Query Match 90.9%; Score 30; DB 12; Length 157;
Best Local Similarity 80.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
Db 86 CDAQC 90
RESULT 25
US-10-124-557-88
Sequence 88, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 172 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-10-124-557-88
Query Match 90.9%; Score 30; DB 12; Length 172;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
Db 86 CDAQC 90
RESULT 26
US-10-124-557-32
Sequence 32, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-124-557-32

Query Match 90.9%; Score 30; DB 12; Length 188;
Best Local Similarity 80.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 CDAEC 5
|||:|
Db 86 CDAQC 90

RESULT 27
US-10-124-557-90
Sequence 90 Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-10-124-557-90

Query Match 90.9%; Score 30; DB 12; Length 192;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CDAEC 5
|||:|
Db 86 CDAQC 90

RESULT 28
US-10-124-557-92
Sequence 92 Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 92:
US-10-124-557-92

Query Match 90.9%; Score 30; DB 12; Length 204;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 CDAEC 5
|||:|
Db 86 CDAQC 90

RESULT 29
US-10-124-557-116
Sequence 116, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 116:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 116:
US-10-124-557-116
Query Match 90.9%; Score 30; DB 12; Length 207;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
Db 86 CDAQC 90
RESULT 30
US-10-124-557-136
Sequence 136, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.

Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 136:
US-10-124-557-136
Query Match 90.9%; Score 30; DB 12; Length 207;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
Db 86 CDAQC 90
RESULT 31
US-10-124-557-132
Sequence 132, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/543,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 208 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 132:
US-10-124-557-132

Query Match 90.9%; Score 30; DB 12; Length 208;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 86 CDAQ 90

RESULT 32
US-10-124-557-94
Sequence 94, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 209 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-10-124-557-94

Query Match 90.9%; Score 30; DB 12; Length 209;
Best Local Similarity 80.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 86 CDAQ 90

RESULT 33
US-10-124-557-96
Sequence 96, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 96:
US-10-124-557-96

Query Match 90.9%; Score 30; DB 12; Length 220;
Best Local Similarity 80.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 34

US-10-124-557-30
Sequence 30, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Query Match 90.9%; Score 30; DB 12; Length 231;
Best Local Similarity 80.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 35

US-10-124-557-72
Sequence 72, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Query Match 90.9%; Score 30; DB 12; Length 237;
Best Local Similarity 80.0%; Pred. No. 3.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 86 CDAQC 90

RESULT 36

US-10-124-557-70
Sequence 70, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Sequence 70, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Gesner, Rodney M.
Hewick, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 70:
US-10-124-557-70
Query Match 90.9%; Score 30; DB 12; Length 296;
Best Local Similarity 80.0%; Pred. No. 4.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-10-124-557-64
Query Match 90.9%; Score 30; DB 12; Length 372;
Best Local Similarity 80.0%; Pred. No. 5.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

TYPE: PRT
ORGANISM: Trichoderma longibrachiatum
US-09-916-494A-14

Query Match 90.9%; Score 30; DB 9; Length 377;
Best Local Similarity 80.0%; Pred. No. 5.8e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 171 CDAQC 175

RESULT 39
US-10-124-557-68
Sequence 68, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesser, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserit, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-124-557-68

Query Match 90.9%; Score 30; DB 12; Length 422;
Best Local Similarity 80.0%; Pred. No. 6.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 45 CDAQC 49

RESULT 40
US-08-841-636A-33
Sequence 33, Application US/08841636A
Patent No. US20020168751A1
GENERAL INFORMATION:

APPLICANT: Mettinen-Oinonen, Arja
APPLICANT: Lomdebrough, John
APPLICANT: Vehmaanger, Jari
APPLICANT: Haakana, Hei
APPLICANT: M nyl, Arja
APPLICANT: Lantto, Ralja
APPLICANT: Elouainio, Minna
APPLICANT: Joutsio, Vesa
APPLICANT: Palohelmo, Maria
APPLICANT: Suominen, Pirkko

TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/841,636A
FILING DATE: 30-APR-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,335
FILING DATE: 17-OCT-1995
APPLICATION NUMBER: US 60/007,926
FILING DATE: 04-DEC-1995
APPLICATION NUMBER: US 60/020,840
FILING DATE: 28-JUN-1996
APPLICATION NUMBER: US 08/732,181
FILING DATE: 16-OCT-1996
APPLICATION NUMBER: PCT/Fin6/00550
FILING DATE: 17-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Timothy J. Shea, Jr.
REGISTRATION NUMBER: 41,306
REFERENCE/DOCKET NUMBER: 1716.0510005/MAC/TUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540

INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Melanocarpus albomyces
STRAIN: ALK04237
FEATURE:

NAME/KEY: Protein
LOCATION: 1..428
OTHER INFORMATION: /label= 50K-cellulase
US-08-841-636A-33

Query Match 90.9%; Score 30; DB 7; Length 428;
 Best Local Similarity 80.0%; Pred. No. 6.4e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 |||:
 Db 189 CDAQC 193

Search completed: July 11, 2003, 15:38:26
 Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 15:19:20 ; Search time 47 Seconds
(without alignments)
10.227 Million cell updates/sec

Title: US-10-006-869-910

Sequence: 1 CDAEC 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	1680	2 A43434	furin (EC 3.4.21.7
2	30	90.9	66	1 NTSR4E	neurotoxin M4 - 1
3	30	90.9	123	2 S40736	hypothetical prote
4	30	90.9	226	2 S50642	hypothetical prote
5	30	90.9	245	2 S43565	ROIHI0.4 protein (
6	30	90.9	411	2 T29475	hypothetical prote
7	30	90.9	423	2 JC7722	cellulase (EC 3.2.
8	30	90.9	451	2 JB0313	exoglucanase (EC 3
9	30	90.9	459	2 A25928	cellulase (EC 3.2.
10	30	90.9	463	2 A48375	cellulase (EC 3.2.
11	30	90.9	464	2 JC7143	endoglucanase I -
12	30	90.9	471	2 F64213	hypothetical prote
13	30	90.9	516	2 S42093	cellulose 1,4-beta
14	30	90.9	525	1 S38794	cellulose 1,4-beta
15	30	90.9	728	2 T20561	hypothetical prote
16	30	90.9	778	2 D95912	probable membrane
17	30	90.9	1429	2 S06434	homeotic protein 1
18	29	87.9	65	1 NTSR3C	neurotoxin 3 - bar
19	29	87.9	104	2 E81283	hypothetical prote
20	29	87.9	147	2 H69180	conserved hypochet
21	29	87.9	160	2 JQ0542	185k secretory pro
22	29	87.9	166	2 F82774	hypothetical prote
23	29	87.9	167	2 F84867	hypothetical prote
24	29	87.9	207	2 S37618	vgr protein - rat
25	29	87.9	267	2 B82694	copper homeostasis
26	29	87.9	307	2 T22128	hypothetical prote
27	29	87.9	326	2 JQ1437	hemagglutinin - in
28	29	87.9	326	2 A70455	beta lactamase pre
29	29	87.9	339	2 A70321	histidine kinase s

30	29	87.9	343	1 HM1VTA	hemagglutinin HA1
31	29	87.9	343	2 JQ2370	hemagglutinin - in
32	29	87.9	343	2 JQ2371	hemagglutinin - in
33	29	87.9	343	2 JQ2372	hemagglutinin - in
34	29	87.9	344	2 J01643	hemagglutinin - in
35	29	87.9	344	2 PQ0319	hemagglutinin - in
36	29	87.9	344	2 PQ0321	hemagglutinin - in
37	29	87.9	345	2 T16573	hypothetical prote
38	29	87.9	345	2 T04016	hypothetical prote
39	29	87.9	353	2 S16785	hemagglutinin - in
40	29	87.9	354	2 A35788	hemagglutinin - in
41	29	87.9	385	2 C87250	dnal protein limpo
42	29	87.9	418	2 T40913	probable integral
43	29	87.9	426	2 JH0690	bone morphogenetic
44	29	87.9	434	2 D83456	probable guanin d
45	29	87.9	447	2 G97067	adenosylmethionine

ALIGNMENTS

RESULT 1
A43434
furin (EC 3.4.21.7) 2 - fruit fly (Drosophila melanogaster).
C:Species: Drosophila melanogaster
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #ext_change 31-Mar-2000
C/Accession: A43434
R:Roebroek, A.J.; Creemers, J.W.; Pauli, I.G.; Kurzik-Dumke, U.; Rentrop, M.; Gateff, E./
U. Biol. Chem. 267, 17208-17215, 1992
A>Title: Cloning and functional expression of Dfurin2, a subtilisin-like proprotein proc
A/Reference number: A43434; PMID:92381036; PMID:1512259
A/Accession: A43434
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1680 <ROR>
A/Cross-references: GB:M94375; NID:G157461; PTD:G157462
A/Note: sequence extracted from NCBI backbone (NCBIN:111933; NCBIR:111934)
C/Genetics:
A/Gene: FlyBase:Fur2
A/Cross-references: FlyBase:FBgn004598
C/Superfamily: subtilisin homology
C/Keywords: hydrolase; serine proteinase; transmembrane protein
F:409-652/Domain: subtilisin homology <SBT>
F:418-457,638/Active site: Asp, His, Ser #status predicted

Query Match
Best Local Similarity 100.0%; Score 33; DB 2; Length 1680;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 968 CDAEC 972

RESULT 2
NTSR4E
neurotoxin M4 - lesser Asian scorpion
C:Species: Mesobuthus eupeus (lesser Asian scorpion)
C>Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #ext_change 31-Dec-1993
C/Accession: J00020
R/Volkova, T.M.; Garsia, A.F.; Telezhinskaya, I.N.; Potapenko, N.A.; Grishin, E.V.
Bioorg. Khim. 11, 1445-1456, 1985
A>Title: Study of neurotoxins from the venom of central Asian scorpion Butkus eupeus.
A/Reference number: A94652; PMID:86130759; PMID:4091860
A/Accession: J00020
A/Molecule type: protein
A/Residues: 1-66 <VOL>
C/Comment: The venom of this scorpion contains three major polypeptide neurotoxins: M9, F,
C/Superfamily: scorpion neurotoxin
C/Keywords: neurotoxin; venom
F:12-65,16-36,22-46,26-48/Diulfide bonds: #status predicted
Query Match 90.9%; Score 30; DB 1; Length 66;

Best Local Similarity 80.0%; Pred. No. 55;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
Db 22 CDSSEC 26

RESULT 3

S40736

hypothetical protein ZK507.2 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Sep-2000

C/Accession: S40736

R/Hawkins, T.; Thomas, K.

Submitted to the EMBL Data Library, December 1993

A/Reference number: S40735

A/Accession: S40736

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-123 <HAM>

A/Cross-references: EMBL:Z29116; NID:G1067087; PID:G439266

C/Genetics:

A/Introns: 24/2; 55/3

C/Superfamily: Caenorhabditis elegans hypothetical protein ZK507.2

Query Match 90.9%; Score 30; DB 2; Length 123;
Best Local Similarity 80.0%; Pred. No. 89;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5

Db 86 CDAQC 90

RESULT 4

S50642

hypothetical protein YER139c - Yeast (Saccharomyces cerevisiae)

C/Species: Saccharomyces cerevisiae

C/Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002

C/Accession: S50642

R/Dietrich, F.S.

Submitted to the EMBL Data Library, December 1994

A/Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda

A/Reference number: S50430

A/Accession: S50642

A/Molecule type: DNA

A/Residues: 1-226 <DBE>

A/Cross-references: EMBL:U18917; NID:G603377; PID:G603379; GSPDB:GN00005; MIPS:YER139C

C/Genetics:

A/Map position: SCD:YER139C

A/Cross-references: SGD:S0000941

A/Map position: 5R

Query Match 90.9%; Score 30; DB 2; Length 226;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5

Db 38 CDSSEC 42

RESULT 5

S43565

ROH10.4 protein (clone ROH10) - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 09-Sep-1997

C/Accession: S43565

R/Lighting, J.

Submitted to the EMBL Data Library, March 1994

A/Reference number: S43563

A/Accession: S43565

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-245 <LIG>
A/Cross-references: EMBL:Z31590; NID:G468568; PID:G468571
C/Genetics:

A/Introns: 52/2; 118/2; 158/2

Query Match 90.9%; Score 30; DB 2; Length 245;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5

Db 205 CDAQC 209

RESULT 6

T29475

hypothetical protein T01D1.6 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C/Accession: T29475

R/Bradshaw, H.; Wohldmann, P.

Submitted to the EMBL Data Library, November 1996

A/Description: The sequence of C. elegans cosmid T01D1.

A/Reference number: Z20623

A/Accession: T29475

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-411

A/Cross-references: EMBL:U80455; PID:AA37887.1; GSPDB:GN00020; CESP:T01D1.6

A/Experimental source: strain Bristol N2; clone T01D1

C/Genetics:

A/Map position: 204/3

A/Introns: 25/3; 304/3

C/Superfamily: gliadin

Query Match 90.9%; Score 30; DB 2; Length 411;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5

Db 337 CDAQC 341

RESULT 7

JC7722

cellulase (EC 3.2.1.4) - fungus (Cochliobolus carbonum)

C/Species: Cochliobolus carbonum

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C/Accession: JC7722

R/Ann, J.H.; Sposato, P.; Kim, S.I.; Walton, J.D.

BioSci. Biotechnol. Biochem. 65, 1406-1411, 2001

A/Title: Molecular cloning and characterization of cel2 from the fungus Cochliobolus carb

A/Reference number: JC7722; NCID:21364349; PMID:11471744

A/Accession: JC7722

A/Molecule type: DNA

A/Residues: 1-423 <AHN>

A/Cross-references: GB:AF336799

A/Experimental source: strain wild-type

C/Comment: This enzyme, a cell wall degrading enzyme, belongs to family 7 of glucanhydrolase

C/Genetics:

A/Map position: 2.5-Mb

A/Introns: 201/3

C/Keywords: glycosidase; hydrolase

F.1.20/Domain: signal sequence #status predicted <SIG>

Query Match 90.9%; Score 30; DB 2; Length 423;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 188 CDAQC 192

RESULT 8

JE0313
exoglucanase (EC 3.2.-.-) - imperfect fungus (*Humicola grisea*)
C/Species: *Humicola grisea*
C/Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C/Accession: JE0313
R/Takashima, S.; Ikura, H.; Nakamura, A.; Hidaka, M.; Masaki, H.; Uozumi, T.
J. Biochem. 124, 717-725, 1998
A/Title: Isolation of the gene and characterization of the enzymatic properties of a major
A/Reference number: JE0313; MUID:98429588; PMID:9756616
A/Accession: JE0313
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-451 <TRK>
A/Note: references: DDBJ:AB003105
A/Keywords: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain homology
A/Keywords: glycosidase; hydrolase

Query Match 90.9%; Score 30; DB 2; Length 451;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 192 CDAQC 196

RESULT 9

A25928
cellulase (EC 3.2.1.4) A precursor - fungus (*Trichoderma reesei*)
N/Alternate names: endo-1,4-beta-glucanase; endoglucanase I
C/Species: *Trichoderma reesei*
C/Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 22-Jun-1999
C/Accession: A25928; A25565
R/Van Arsdell, J.N.; Kwok, S.; Schweickart, V.L.; Laderer, M.B.; Gelfand, D.H.; Innis, M.
Bio/Technology 5, 60-64, 1987
A/Title: Cloning, characterization, and expression in *Saccharomyces cerevisiae* of endogl
A/Reference number: A25928
A/Accession: A25928
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-459 <VAN>
A/Experimental source: strain L27
A/Title: Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide seq
A/Reference number: A25565; MUID:87106822; PMID:2948877
A/Accession: A25565
A/Molecule type: mRNA
A/Residues: 1-459 <PEN>
A/Cross-references: GB:115665; NID:9170546; PION:AAA4212.1; PID:9170547
C/Genetics:

A/Status: eg11
A/Intons: 257/2; 454/1
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A/pathway: cellulose degradation
C/Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain hom
C/Keywords: glycosidase; hydrolase; polysaccharide degradation
F/1-22/Domain: signal sequence #status predicted <SIG>
F/23-459/Product: cellulase A #status predicted <MAT>
F/428-459/Domain: fungal cellulose-binding domain homology <FCB>
Query Match 90.9%; Score 30; DB 2; Length 459;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 193 CDAQC 197

RESULT 10

A48375
cellulase (EC 3.2.1.4) - fungus (*Trichoderma longibrachiatum*)
N/Alternate names: endo-1,4-beta-glucanase; endoglucanase I homolog EGL1
C/Species: *Trichoderma longibrachiatum*
C/Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C/Accession: A48375; S28521
R/Gonzalez, R.; Ramon, D.; Perez-Gonzalez, J.A.
Appl. Microbiol. Biotechnol. 38, 370-375, 1992
A/Title: Cloning, sequence analysis and yeast expression of the eg11 gene from *Trichodem*
A/Reference number: A48375; MUID:93159747; PMID:1369161
A/Contents: CECT 2606
A/Accession: A48375
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-463 <GON>
A/Cross-references: EMBL:X60652; NID:95181; PION:CAA43059.1; PID:95182
A/Note: sequence extracted from NCBI backbone (NCBIN:125157, NCBIR:125158)
C/Genetics:
A/Intons: 257/2; 458/1
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A/pathway: cellulose degradation
C/Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain homo
C/Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F/432-463/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 90.9%; Score 30; DB 2; Length 463;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 193 CDAQC 197

RESULT 11

JC7143
endoglucanase I - fungus (*Trichoderma viride*)
C/Species: *Trichoderma viride*
C/Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 03-Jun-2000
C/Accession: JC7143
R/Kwon, I.; Ekino, K.; Goto, M.; Furukawa, K.
Biosci. Biotechnol. Biochem. 63, 1714-1720, 1999
A/Title: Heterologous expression and characterization of endoglucanase I (Egl) from *Trich*
A/Reference number: JC7143; MUID:20052947; PMID:10586500
A/Accession: JC7143
A/Molecule type: mRNA
A/Residues: 1-464 <KMO>
A/Experimental source: HK-75
A/Comment: This protein is a fusion glycoprotein with catalytic and cellulose binding dom
C/Genetics:

A/Status: eg11
A/Intons: 257/2; 459/1
C/Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain homo
C/Keywords: fusion protein; glycoprotein
F/433-464/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 90.9%; Score 30; DB 2; Length 464;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|||:
Db 193 CDAQC 197

RESULT 12

F64213
 hypothetical protein homolog MG123 - Mycoplasma genitalium
 C/Species: Mycoplasma genitalium
 C/Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 07-Dec-1999
 C/Accession: F64213
 R/Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
 M.; Fuhrman, J.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
 C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A/Title: The minimal gene complement of Mycoplasma genitalium.
 A/Reference number: A64200; MUID:96026346; PMID:7559993
 A/Accession: F64213
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-471 <TRIG>
 A/Cross-references: GB:U39691; GB:L43967; NID:g1045794; PID:g1045803; TIGR:MG123
 A/Experimental source: strain G-37
 C/Genetics:
 A/Genetic code: SGC3
 Superfamily: hypothetical protein MG123

Query Match 90.9%; Score 30; DB 2; Length 471;
 Best Local Similarity 80.0%; Pred. No. 2.5e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 191 CDADC 195

RESULT 13
 S42093
 cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Neurospora crassa
 C/Species: Neurospora crassa
 C/Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C/Accession: S42093
 R/Taleb, F.; Radford, A.
 submitted to the EMBL Data Library, February 1994
 A/Description: Cloning sequencing and homologies of the CBH-1 (exocellobiohydrolase) gene
 A/Reference number: S42093
 A/Accession: S42093
 A/Molecule type: DNA
 A/Residues: 1-516 <TL>
 A/Cross-references: EMBL:X77778; NID:g456657; PID:CAA54815.1; PID:g456658
 C/Genetics:
 A/Intons: 227/3
 Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homologs
 A/Keywords: glycosidase; hydrolase; polysaccharide degradation
 A/Map position: 35-516/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 90.9%; Score 30; DB 2; Length 516;
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 184 CDAQC 188

RESULT 14
 S38794
 cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (Humicola grisea)
 N/Alternate names: beta-glucanocellobiohydrolase; exoglucanase
 C/Species: Humicola grisea var. thermidea
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: S38794; S08240; A45869
 R/Radford, A.
 submitted to the EMBL Data Library, June 1991
 A/Reference number: S38794
 A/Accession: S38794
 A/Molecule type: DNA
 A/Residues: 1-525 <RAD>
 A/Cross-references: EMBL:X17258; NID:g2760; PID:CAA5159.1; PID:g2761

A/Note: this is a revision to the sequence from reference S08240
 R/de Oliveira Azevedo, M.; Radford, A.
 Nucleic Acids Res. 18, 668, 1990
 A/Title: Sequence of cbh-1 gene of Humicola grisea var. thermidea.
 A/Reference number: S08240; NUID:90175006; PMID:2208855
 A/Accession: S08240
 A/Molecule type: DNA
 A/Residues: 1-299, 'H', 301-525 <DEO>
 A/Cross-references: EMBL:X17258
 A/Note: the authors translated the codon CAG for residue 87 as His
 A/Note: this sequence has been revised in reference S38794
 R/Azevedo, M.; de O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.
 J. Gen. Microbiol. 136, 2569-2576, 1990
 A/Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of Humicola
 A/Reference number: A45869; NUID:91178527; PMID:2127803
 A/Accession: A45869
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-20, 'R', 22-34, 'K', 35-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QQH', 241-244, 'I'
 A/Cross-references: GB:M64588; GB:X17258
 A/Note: this sequence has been revised. See entry S08240
 C/Genetics:
 A/Intons: 138/1
 Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homologs
 A/Keywords: glycosidase; hydrolase; polysaccharide degradation
 F/494-525/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 90.9%; Score 30; DB 1; Length 525;
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 191 CDADC 195

RESULT 15
 T20561
 hypothetical protein F07H5.8 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: T20561
 R/Steward, C.
 submitted to the EMBL Data Library, December 1995
 A/Reference number: Z19292
 A/Accession: T20561
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-728 <WIL>
 A/Cross-references: EMBL:Z68314; PIDN:CAA92662.1; GSPDB:GN00020; CESP:F07H5.8
 A/Experimental source: clone F07H5
 C/Genetics:
 A/Intons: 32/3; 81/3; 489/3; 591/3; 633/3
 A/Map position: 2
 A/Intons: 32/3; 81/3; 489/3; 591/3; 633/3

Query Match 90.9%; Score 30; DB 2; Length 728;
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 99 CDAQC 103

RESULT 16
 D95912
 probable membrane protein, slightly similar to carrier proteins [imported] - Sinorhizobium
 C/Species: Sinorhizobium meliloti
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: D95912
 R/Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernand

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A/Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A/Reference number: A95842; MWID:21396508; PMID:11481431
 A/Accession: D95912
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-778 <GUR>
 A/Cross-references: GB:AL591985; PIDN:CAC48964.1; PID:G15140449; GSPDB:GN00167
 A/Experimental source: strain 1021, megaplasmid pSymB
 R/Galbert, F.; Finan, T.M.; Long, S.R.; Publer, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davies, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A/Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A/Reference number: A96039; MWID:21368234; PMID:11474104
 A/Contents: annotation
 C/Genetics:
 A/Genes: SMD20824
 A/Genome: plasmid

Query Match 90.9%; Score 30; DB 2; Length 778;
 Best Local Similarity 80.0%; Pred. No. 3.7e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CDABC 5
 134 CDADC 138

RESULT 17

S06434
 homeotic protein lin-12 precursor - *Caenorhabditis elegans*
 C/Species: *Caenorhabditis elegans*
 C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 20-Sep-1999
 C/Accession: S06434; A24769
 R/Yochem, J.; Weston, K.; Greenwald, I.
 Nature 335, 547-550, 1988

A/Title: The *Caenorhabditis elegans* lin-12 gene encodes a transmembrane protein with ove
 A/Reference number: S06434; MWID:8834747; PMID:3419531
 A/Accession: S06434

A/Molecule type: DNA
 A/Residues: 1-1429 <YOC>
 A/Cross-references: EMBL:M12069; NID:G156357; PIDN:AAH70191.1; PID:G156358

R/Greenwald, I.
 Cell 43, 583-590, 1985

A/Reference number: A24769; MWID:86079540; PMID:3000611
 A/Accession: A24769

A/Molecule type: DNA
 A/Residues: 173-712 <GR>

C/Genetics:
 A/Insertions: 50/2, 90/1, 109/1, 172/3, 545/1, 589/2, 632/2, 1273/3, 1389/3
 C/Superfamily: unassigned ankylrin repeat proteins; ankylrin repeat homology; EGF homology
 C/Keywords: glycoprotein; transmembrane protein
 F/254-284/Domain: EGF homology <EGF1>
 F/507-540/Domain: EGF homology <EGF2>
 F/547-578/Domain: EGF homology <EGF2>
 F/909-931/Domain: transmembrane #status predicted, <TMM>
 F/1093-1125/Domain: ankylrin repeat homology <AN1>
 F/1206-1238/Domain: ankylrin repeat homology <AN2>
 F/1240-1272/Domain: ankylrin repeat homology <AN3>

Query Match 90.9%; Score 30; DB 2; Length 1429;
 Best Local Similarity 80.0%; Pred. No. 5.9e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CDABC 5
 652 CDADC 656

RESULT 18

NTSR3C
 neurotoxin 3 - bark scorpion
 C/Species: *Centruroides sculpturatus* (bark scorpion)
 C/Date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 23-Aug-1996
 C/Accession: A90058; A94470; A01754
 R/Abidin, D.R.; Watt, D.D.; Goos, S.M.; Mlejnek, R.V.
 Arch. Biochem. Biophys. 164, 694-706, 1974
 A/Title: Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides*
 A/Reference number: A90058; MWID:75163395; PMID:4460885
 A/Accession: A90058
 A/Molecule type: protein
 A/Residues: 1-24, 'NTC', 28-63, 'CS' <BAB>
 R/Shom, A.; Moile, J.
 unpublished results, cited by Fontecilla-Camps, J.C., et al., *Toxicon* 20, 1-7, 1982
 A/Reference number: A94470
 A/Accession: A94470
 A/Molecule type: protein
 A/Residues: 1-65 <BHO>
 R/Fontecilla-Camps, J.C.; Almassy, R.J.; Suddath, F.L.; Buga, C.E.
Toxicon 20, 1-7, 1982
 A/Title: The three-dimensional structure of scorpion neurotoxins.
 A/Reference number: A94314; MWID:82200153; PMID:7080025
 A/Contents: annotation; X-ray crystallography, 1.8 angstroms; disulfide bonds
 A/Note: X-ray crystallographic studies were based on the revised sequence shown
 C/Superfamily: scorpion neurotoxin
 C/Keywords: neurotoxin; venom
 F/12-65,16-41,25-46,29-48/Disulfide bonds: #status experimental

Query Match 87.9%; Score 29; DB 1; Length 65;
 Best Local Similarity 80.0%; Pred. No. 83;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CDABC 5
 25 CDTEC 29

RESULT 19

E81283
 hypothetical protein Cj1384c (imported) - *Campylobacter jejuni* (strain NCTC 11168)
 C/Species: *Campylobacter jejuni*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C/Accession: E81283

R/Farkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett
 Nature 403, 665-668, 2000

A/Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp
 A/Reference number: A81250; MWID:20150912; PMID:10688204
 A/Accession: E81283

A/Status: preliminary
 A/Molecule type: DNA

A/Residues: 1-104 <PAR>
 A/Cross-references: GB:AL139078; GB:AL111168; NID:G6968723; PIDN:CAB73811.1; PID:G6968814

A/Experimental source: serotype O2, strain NCTC 11168
 C/Genetics:

A/Genes: Cj1384c
 C/Superfamily: *Campylobacter jejuni* hypothetical protein Cj1384c

Query Match 87.9%; Score 29; DB 2; Length 104;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CDABC 5
 44 CDTEC 48

RESULT 20

H69180
 conserved hypothetical protein MTH609 - *Methanobacterium thermoautotrophicum* (strain Delt
 C/Species: *Methanobacterium thermoautotrophicum*
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Oct-1999
 C/Accession: H69180

R.Smith, D.R.: Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadator, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N. J. Bacteriol. 179, 7135-7155, 1997

A>Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: funct

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: H69180

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-147 <MTH>

A:Cross-references: GB:AE000842; GB:AE000666; NID:g2621676; PIDN:AA85115.1; PID:g262168

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH609

A:Start codon: TTG

C:Superfamily: *Methanococcus jannaschii* conserved hypothetical protein MJ1627

Query Match 87.9%; Score 29; DB 2; Length 147;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 96 CDVEC 100

RESULT 21

J00542

185K secretory protein - midge (*Chironomus tentans*) (fragment)

N:Alternate names: balbiani ring 3 protein

C:Species: *Chironomus tentans*

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 17-Mar-2000

C:Accession: J00542

R:Dignam, S.S.; Case, S.T. Gene 88, 133-140, 1990

A>Title: Balbiani ring 3 in *Chironomus tentans* encodes a 185-kDa secretory protein which

A:Reference number: J00542; MUID:90269600; PMID:2189782

A:Accession: J00542

A:Molecule type: mRNA

A:Residues: 1-160 <DIG>

A:Cross-references: GB:M24160

A:Experimental source: salivary gland

C:Superfamily: unassigned Balbiani ring proteins

Query Match 87.9%; Score 29; DB 2; Length 160;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 148 CDVEC 152

RESULT 22

F82774

hypothetical protein XF0681 [imported] - *Xylella fastidiosa* (strain 945c)

C:Species: *Xylella fastidiosa*

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: F82774

R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20355717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: F82774

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-166 <SIM>

A:Cross-references: GB:AE003912; GB:AE003849; NID:g9105560; PIDN:AAF83491.1; GSPDB:GN001

A:Experimental source: strain 945c

R:Simmons, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

es-Neco, E.; Docena, C.; El-Dorcy, H.; Facincani, A.P.; Ferreira, A.U.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme

J.D.; Junqueira, M.L.; Kemp, E.L.; Kiteajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laigre

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins E.

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Ze

A:Reference number: A59328

A:Accession: S37618

A:Contents: annotation

C:Genetics:

A:Gene: XF0681

Query Match 87.9%; Score 29; DB 2; Length 166;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 153 CDVEC 157

RESULT 23

F84867

hypothetical protein At2g43560 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: F84867

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: F84867

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-167 <STO>

A:Cross-references: GB:AB002093; NID:g2289010; PIDN:AA84339.1; GSPDB:GN00139

A:Genetics:

A:Gene: At2g43560

A:Map position: 2

Query Match 87.9%; Score 29; DB 2; Length 167;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 32 CEVEC 36

RESULT 24

S37618

vgr protein - rat (fragment)

C:Species: *Rattus norvegicus* (Norway rat)

C>Date: 18-Feb-1994 #sequence_revision 26-May-1995 #text_change 16-Jul-1999

C:Accession: S37618

R:Sausermann, U.; Meyerman, R.; Schluesener, H.U.

J. Neurosci. Res. 33, 142-147, 1992

A>Title: Cloning of a novel TGF-beta related cytokine, the vgr, from rat brain: cloning c

A:Reference number: EMBL:X58630; NID:g57475; PIDN:CAA41634.1; PID:g57476

A:Accession: S37618

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-207 <SAU>

A:Cross-references: EMBL:X58630; NID:g57475; PIDN:CAA41634.1; PID:g57476

A>Note: the sequence from Fig. 4 is inconsistent with that from Fig. 1 in having 88-Ser e

C:Superfamily: inhibin

Query Match 87.9%; Score 29; DB 2; Length 207;
Best Local Similarity 80.0%; Pred. No. 2e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CDABC 5
Db 135 CDVEC 139

RESULT 25

B82694

copper homeostasis protein XF1341 (imported) - Xylella fastidiosa (strain 9asc)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C/Accession: B82694

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A/Reference number: A82515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: B82694

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-267 <STM>

A/Cross-references: GB:AE003966; GB:AE003849; NID:9106327; PIDN:AAF84150.1; GSPDB:GN001

A/Experimental source: strain 9asc

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Birones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carrer, D.M.; Carver, H

as-Neto, E.; Docena, C.; El-Dorry, H.; Facinani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitejima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chao, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E

A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, R.C.; Miyaki, C.Y.

Rodrigues, V.; Rosa, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

A/Authors: da Silva, A.C.R.; da Silva, F.R.; de Sa, R.G.; Santelli, R.V.; Sawasak

M.; Teshako, M.H.; Valada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A/Reference number: A59328

A/Contents: annotation

C/Genetics:

A/Gene: XF1341

Query Match 87.9%; Score 29; DB 2; Length 267;

Best Local Similarity 80.0%; Pred. No. 2.5e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5

Db 101 CDVEC 105

LT 26

T22128

hypothetical protein F43D9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T22128

R:Wilkinson, J.

submitted to the EMBL Data Library, August 1994

A/Reference number: Z19521

A/Accession: T22128

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 1-307 <MTL>

A/Cross-references: EMBL:Z35640; PIDN:CAA84705.1; GSPDB:GN00021; CESP:F43D9.2

C/Experimental source: clone F43D9

C/Genetics:

A/Gene: CESP:F43D9.2

A/Map position: 3

A/Introns: 34/3; 102/3; 185/2; 233/3

Query Match 87.9%; Score 29; DB 2; Length 307;

Best Local Similarity 80.0%; Pred. No. 2.7e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5
Db 219 CDVEC 223

RESULT 27

J01437

hemagglutinin - influenza A virus (strain A/NTB/4/88 (H1N1))

C:Species: Influenza A virus

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999

C/Accession: J01437

R:Robertson, J.S.; Nicolson, C.; Bootman, J.S.; Major, D.; Robertson, E.W.; Wood, J.M.

J. Gen. Virol. 72, 2671-2677, 1991

A>Title: Sequence analysis of the haemagglutinin (HA) of influenza A (H1N1) viruses prese

A/Reference number: J01437; MUID:92044437; PMID:1940864

A/Accession: J01437

A/Molecule type: genomic RNA

A/Residues: 1-326 <ROB>

A/Cross-references: EMBL:X59778

A/Status: preliminary

C/Keywords: hemagglutinin; homotrimer

Query Match 87.9%; Score 29; DB 2; Length 326;

Best Local Similarity 80.0%; Pred. No. 2.9e+02;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDABC 5

Db 275 CDKAC 279

RESULT 28

A70455

beta lactamase precursor - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000

C/Accession: A70455

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

A/Reference number: A70455

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-326 <AOFS>

A/Cross-references: GB:AE000757; NID:g2984092; PIDN:AAC07632.1; PID:g2984094; GB:AE00065;

A/Experimental source: strain VFS

C/Genetics:

A/Gene: cpH1

C/Superfamily: Streptomyces actinorhodin polyketide synthase system probable dehydrase

Query Match 87.9%; Score 29; DB 2; Length 326;

Best Local Similarity 80.0%; Pred. No. 2.9e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5

Db 277 CDVEC 281

RESULT 29

A70321

histidine kinase sensor protein - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Aug-2002

C/Accession: A70321

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-326 <AOFS>

A/Cross-references: GB:AE000757; NID:g2984092; PIDN:AAC07632.1; PID:g2984094; GB:AE00065;

A/Experimental source: strain VFS

C/Genetics:

A/Gene: cpH1

C/Superfamily: Streptomyces actinorhodin polyketide synthase system probable dehydrase

Query Match 87.9%; Score 29; DB 2; Length 326;

Best Local Similarity 80.0%; Pred. No. 2.9e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5

Db 277 CDVEC 281

RESULT 29

A70321

histidine kinase sensor protein - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Aug-2002

C/Accession: A70321

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-326 <AOFS>

A/Cross-references: GB:AE000757; NID:g2984092; PIDN:AAC07632.1; PID:g2984094; GB:AE00065;

A/Experimental source: strain VFS

C/Genetics:

A/Gene: cpH1

C/Superfamily: Streptomyces actinorhodin polyketide synthase system probable dehydrase

Query Match 87.9%; Score 29; DB 2; Length 326;

Best Local Similarity 80.0%; Pred. No. 2.9e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5

Db 277 CDVEC 281

A/Accession: A70321
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-339 <AQP>
A/Cross-references: GB:AE000679; NID:92982936; PIDN:AA06560.1; PID:92982944; GB:AE00065
A/Experimental source: strain VFS
C/Genetics:
A/Gene: hkeP4
C/Superfamily: histidine kinase sensor protein; sensor histidine kinase homology
F/112-337/Domain: sensor histidine kinase homology <SHK>

Query Match 87.9%; Score 29; DB 2; Length 339;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDARC 5
DB 58 CDGEC 62

RESULT 30
HMITVA
hemagglutinin HA1 - influenza A virus (strain A/Taiwan/1/86 [HA1])
C/Species: Influenza A virus
C/Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Apr-1994
C/Accession: A26765
R/Robertson, J.S.
J. Gen. Virol. 68, 1205-1208, 1987
A/Title: Sequence analysis of the haemagglutinin of A/Taiwan/1/86, a new variant of human
A/Reference number: A26765; MUID:87197215; PMID:3572359
A/Accession type: DNA
A/Molecule type: DNA
A/Residues: 1-343 <ROB>
C/Genetics:
A/Map position: segment 4
C/Superfamily: Influenza virus hemagglutinin
C/Keywords: glycoprotein; hemagglutinin; homotrimer
F/27,28,40,71,104,142,172,177,286,304/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 87.9%; Score 29; DB 1; Length 343;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDARC 5
DB 292 CDARC 296

RESULT 31
JQ2370
hemagglutinin - influenza A virus (strain SN0690)
C/Species: Influenza A virus
C/Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 07-May-1999
C/Accession: JQ2370; JQ2369
R/Rocha, E.P.; Xu, X.; Hall, H.E.; Allen, J.R.; Regnery, H.L.; Cox, N.J.
J. Gen. Virol. 74, 2513-2518, 1993
A/Title: Comparison of 10 influenza A (H1N1 and H3N2) haemagglutinin sequences obtained
A/Reference number: JQ2369; MUID:94065682; PMID:8245870
A/Accession: JQ2370
A/Molecule type: mRNA
A/Residues: 1-343 <ROCI>
A/Cross-references: GB:L20109
A/Experimental source: subtype H1N1, strain SN0690
A/Accession: JQ2369
A/Molecule type: mRNA
A/Residues: 1-343 <ROCC>
A/Cross-references: GB:L20107
A/Experimental source: subtype H1N1, strain SN0390
C/Superfamily: influenza virus hemagglutinin

Query Match 87.9%; Score 29; DB 2; Length 343;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDARC 5
DB 292 CDARC 296

RESULT 32
JQ2371
hemagglutinin - influenza A virus (strain SN1090)
C/Species: Influenza A virus
C/Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 07-May-1999
C/Accession: JQ2371
R/Rocha, E.P.; Xu, X.; Hall, H.E.; Allen, J.R.; Regnery, H.L.; Cox, N.J.
J. Gen. Virol. 74, 2513-2518, 1993
A/Title: Comparison of 10 influenza A (H1N1 and H3N2) haemagglutinin sequences obtained
A/Reference number: JQ2369; MUID:94065682; PMID:8245870
A/Accession: JQ2371
A/Molecule type: mRNA
A/Residues: 1-343 <ROCC>
A/Cross-references: GB:L20111
A/Experimental source: subtype H1N1
C/Superfamily: Influenza virus hemagglutinin

Query Match 87.9%; Score 29; DB 2; Length 343;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDARC 5
DB 292 CDARC 296

RESULT 33
JQ2372
hemagglutinin - influenza A virus (strain SN1190 and SN1290)
C/Species: Influenza A virus
C/Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 07-May-1999
C/Accession: JQ2372
R/Rocha, E.P.; Xu, X.; Hall, H.E.; Allen, J.R.; Regnery, H.L.; Cox, N.J.
J. Gen. Virol. 74, 2513-2518, 1993
A/Title: Comparison of 10 influenza A (H1N1 and H3N2) haemagglutinin sequences obtained
A/Reference number: JQ2369; MUID:94065682; PMID:8245870
A/Accession: JQ2372
A/Molecule type: mRNA
A/Residues: 1-343 <ROCC>
A/Cross-references: GB:L20113
A/Experimental source: subtype H1N1
C/Superfamily: Influenza virus hemagglutinin

Query Match 87.9%; Score 29; DB 2; Length 343;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDARC 5
DB 292 CDARC 296

RESULT 34
JQ1643
hemagglutinin - influenza A virus (strain A/Hebei/24/89 [H2N2]) (fragment)
C/Species: Influenza A virus
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C/Accession: JQ1643
R/Li, X.S.; Zhao, C.Y.; Gao, H.W.; Zhang, Y.Q.; Ishida, M.; Kanegae, Y.; Endo, A.; Nerome
J. Gen. Virol. 73, 1329-1337, 1992
A/Title: Origin and evolutionary characteristics of antigenic reassortant influenza A (H1
A/Reference number: P00408; MUID:92300326; PMID:1607856
A/Accession: JQ1643
A/Molecule type: genomic RNA
A/Residues: 1-344 <LHA>
A/Cross-references: DDBJ:DI0163
C/Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer

Query Match 87.9%; Score 29; DB 2; Length 344;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 292 CDAKC 296

RESULT 35

PQ0319
hemagglutinin - influenza A virus (strains Harbin/1/88 [H1N2] and Harbin/1/89 [H1N2])
C:Species: influenza A virus
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 20-Sep-1999
C:Accession: PQ0319
R:Guo, Y.; Xu, X.; Cox, N.D.
Gen. Virol. 73, 383-388, 1992
Title: Human influenza A (H1N2) viruses isolated from China.
Reference number: PQ0319; PMID:9216746; PMID:1538194

A:Accession: PQ0319
A:Status: nucleic acid sequence not shown
A:Molecule type: genomic RNA
A:Residues: 1-344 <GUO>
A:Cross-references: GB:LI9006; NID:g305144; PIDN:AAA43175.1; PID:g305145
C:Genetics:
A:Gene: HA
C:Superfamily: influenza virus hemagglutinin
C:Keywords: homotrimer

Query Match 87.9%; Score 29; DB 2; Length 344;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 292 CDAKC 296

RESULT 36

PQ0321
hemagglutinin - influenza A virus (strain Xianfeng/3/89 [H1N2]) (fragment)
C:Species: influenza A virus
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 20-Sep-1999
C:Accession: PQ0321
Guo, Y.; Xu, X.; Cox, N.D.
Gen. Virol. 73, 383-388, 1992
Title: Human influenza A (H1N2) viruses isolated from China.
Reference number: PQ0319; PMID:9216746; PMID:1538194

A:Accession: PQ0321
A:Status: nucleic acid sequence not shown
A:Molecule type: genomic RNA
A:Residues: 1-344 <GUO>
A:Cross-references: GB:LI9005; NID:g305195; PIDN:AAA99877.1; PID:g305196
C:Genetics:
A:Gene: HA
C:Superfamily: influenza virus hemagglutinin
C:Keywords: homotrimer

Query Match 87.9%; Score 29; DB 2; Length 344;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 292 CDAKC 296

RESULT 37

T16573

hypothetical protein K05F1.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16573
R:Mohdamm, P.

submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid K05F1.
A:Reference number: Z18537
A:Accession: T16573
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-345 <MOH>
A:Cross-references: EMBL:U29377; NID:g868173; PID:g868181; PIDN:AAA68718.1; CESP:K05F1.1
C:Genetics:
A:Gene: CESP:K05F1.1
A:Introns: 34/2; 39/1; 70/3; 104/3; 138/1; 191/1; 258/1; 300/3

Query Match 87.9%; Score 29; DB 2; Length 345;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 209 CDAKC 213

RESULT 38

T04016
hypothetical protein F17A8.40 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184
A:Accession: T04016
A:Molecule type: DNA
A:Residues: 1-345 <BEV>
A:Cross-references: EMBL:AL049482
A:Experimental source: cultivar Columbia; BAC clone F17A8
C:Genetics:
A:Map position: 4
A:Introns: 271/3
A:Note: F17A8.40

Query Match 87.9%; Score 29; DB 2; Length 345;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 139 CDAKC 143

RESULT 39

S16785
hemagglutinin - influenza A virus
C:Species: influenza A virus
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Sep-1999
C:Accession: S16785
R:Robertson, J.S.; Nicolson, C.; Bootman, J.S.; Major, D.; Robertson, E.W.; Wood, J.M.
submitted to the EMBL Data Library, August 1991
A:Description: Sequence analysis of the HA of influenza A(H1N1) viruses present in clinic
A:Reference number: S16785
A:Accession: S16785
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-353 <ROB>
A:Cross-references: EMBL:X59778; NID:g59294; PIDN:CAA42444.1; PID:g59295
C:Superfamily: influenza virus hemagglutinin

Query Match 87.9%; Score 29; DB 2; Length 353;

Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CDAC 5
|||:|

Db 285 CDAC 289

RESULT 40

A35788 hemagglutinin - influenza A virus (strain A/SL/2/87)

C:Species: influenza A virus

C:Date: 15-Feb-1991 #sequence_revision 03-May-1994 #text_change 03-May-1994

C:Accession: A35788; A37265

R:Rajakumar, A.; Swierkosz, E.M.; Schulze, I.T.

Proc. Natl. Acad. Sci. U.S.A. 87, 4154-4158, 1990

A:Title: Sequence of an influenza virus hemagglutinin determined directly from a clinical

A:Reference number: A35788; MUID:90272672; PMID:2349225

Accession: A35788

Status: preliminary

Molecule type: DNA

A:Residues: 1-354 <RAJ>

C:Cross-references: GB:M33748

C:Superfamily: influenza virus hemagglutinin

C:Keywords: homotrimer

Query Match 87.9%; Score 29; DB 2; Length 354;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CDAC 5
|||:|

Db 292 CDAC 296

Search completed: July 11, 2003, 15:29:18
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 15:24:50 ; Search time 23 Seconds

(without alignments)
9.017 Million cell updates/sec

Title: US-10-006-869-910

Perfect score: 33

Sequence: 1 CDAEC 5

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Maximum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	1680	1	FUR2_DROME
2	30	90.9	66	1	SCXE_BUTEU
3	30	90.9	226	1	YEM9_YEAST
4	30	90.9	331	1	YOO1_CABEL
5	30	90.9	373	1	YOO3_CABEL
6	30	90.9	402	1	GUNI_HUMIN
7	30	90.9	429	1	GUNC_FUSOX
8	30	90.9	435	1	GUNI_HUMGR
9	30	90.9	459	1	GUNI_TIRIE
10	30	90.9	463	1	GUNI_TIRIO
11	30	90.9	471	1	Y123_MYCSE
12	30	90.9	514	1	GUXC_FUSOX
13	30	90.9	516	1	GUX1_NEUCR
14	30	90.9	525	1	GUX1_HUMGR
15	30	90.9	528	1	PPGA_HUMAN
16	30	90.9	1429	1	L112_CABEL
17	29	87.9	61	1	SIX5_ANDAU
18	29	87.9	68	1	ITX1_TEGAG
19	29	87.9	68	1	ITX2_TEGAG
20	29	87.9	68	1	ITX3_TEGAG
21	29	87.9	85	1	SC11_MESMA
22	29	87.9	87	1	SCX1_CENSC
23	29	87.9	147	1	Y609_MERTH
24	29	87.9	207	1	EMPE_RAT
25	29	87.9	223	1	FKB2_ARATH
26	29	87.9	307	1	FKB3_CABEL
27	29	87.9	344	1	HEMA_IHAR
28	29	87.9	344	1	HEMA_IATAI
29	29	87.9	344	1	HEMA_IATXA
30	29	87.9	365	1	DNAJ_CACR
31	29	87.9	395	1	UNIV_STRPU
32	29	87.9	426	1	BMP7_XENIA
33	29	87.9	452	1	BMP5_MOUSE

34	29	87.9	454	1	BMP5_HUMAN
35	29	87.9	510	1	EMPE_MOUSE
36	29	87.9	513	1	BMP6_HUMAN
37	29	87.9	663	1	TRA_BPMU
38	29	87.9	905	1	YAG1_SCHPO
39	29	87.9	1700	1	BAR3_CHITE
40	28	84.8	61	1	SIX2_BUTAR
41	28	84.8	61	1	SIX4_BUTOC
42	28	84.8	61	1	SIX5_BUTOC
43	28	84.8	65	1	SCBK_MESMA
44	28	84.8	65	1	SCXA_BUTEU
45	28	84.8	66	1	SCX4_CENSC

ALIGNMENTS

RESULT 1
FUR2_DROME STANDARD; PRT; 1680 AA.
ID AC P30432;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).
GN FUR2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RX MEDLINE=92381036; PubMed=1512259;
RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,
RA Kentrop M., Gateff E.A.F., Leunissen J.A.M., van de Ven W.J.;
RT Cloning and functional expression of Dfurnin2, a subtilisin-like
RT proprotein processing enzyme of Drosophila melanogaster with multiple
RT repeats of a cysteine motif."
RL J. Biol. Chem. 267:17208-17215(1992).
CC -I- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTASE
CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: Release of mature proteins from their
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg-I-Zaa bonds, where Xaa
CC can be any amino acid and Yaa is Arg or Lys. Releases albumin,
CC complement component C3 and von Willebrand factor from their
CC respective precursors.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M94375; AAA28551.1; -
DR PIR: A43434; A43434.
DR HSSP: Q99405; IMPR.
DR MEROPS: S08_049; -
DR FLYBASE: FBgn0004598; Fur2.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002884; P domain.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF000082; Peptidase_S8; 1.
DR Pfam: PF01483; P; PARTIAL.
DR PRINTS: PR00723; SUBTILISIN.
DR PRODOM: PD000717; P domain; 1.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00261; FU; 10.

DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolyase; Serine protease; Glycoprotein; Signal; Transmembrane;
 KM Multigene family; Zymogen; Repeat.
 FT SIGNAL 1 319 POTENTIAL.
 FT PROPEP ? 319 POTENTIAL.
 FT CHAIN 320 1680 FURIN-LIKE PROTEASE 2.
 FT ACT_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 457 457 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 638 638 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DOMAIN 962 1007 10 X TANDEM REPEATS, CYS-RICH.
 FT REPEAT 962 1007 1.
 FT REPEAT 1008 1057 2.
 FT REPEAT 1058 1104 3.
 FT REPEAT 1105 1153 4.
 FT REPEAT 1154 1205 5.
 FT REPEAT 1206 1254 6.
 FT REPEAT 1255 1299 7.
 FT REPEAT 1300 1346 8.
 FT REPEAT 1347 1393 9.
 FT REPEAT 1394 1444 10.
 FT TRANSMEM 1508 1532 POTENTIAL.
 FT DOMAIN 1533 1680 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 928 928 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1278 1278 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1440 1440 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 1680 AA; 183599 MW; 0A99CEB70A8E293 CRC64;
 Query Match 100.0%; Score 33; DB 1; Length 1680;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 968 CDAEC 972
 ID SCHE BUTEU STANDARD; PRT; 66 AA.
 AC P09982;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurotoxin M14.
 OS Butus eupeus (lesser Asian scorpion) (Mesobuthus eupeus).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
 OC Butidae; Butidae; Mesobuthus.
 OC NCBI_TaxId=34648;
 RX MEDLINE=5046632; PubMed=6497916;
 RA Volkova T.M., Garsia A.F., Telezhinskaya I.N., Potapenko N.A.,
 RA Grishin E.V.;
 RT "amino acid sequence of 2 neurotoxins from the scorpion Butus eupeus
 venom";
 RL Bioorg. Khim. 10:979-982(1984).
 CC -FUNCTION: BINDS TO SODIUM CHANNELS AND INHIBITS THE INACTIVATION
 OF THE ACTIVATED CHANNELS, THEREBY BLOCKING NEURONAL TRANSMISSION.
 CC HAS PARALYTIC ACTIVITY IN MICE.
 CC -SUBCELLULAR LOCATION: Secreted.
 CC -SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.
 CC ALPHA-TOXIN SUBFAMILY.

DR PIR; J00020; NTSRAE.
 DR HSP; P45697; ISN1.
 DR InterPro; IPR003614; Knoc1.
 DR InterPro; IPR001219; Neurotoxin.
 DR InterPro; IPR002061; Scorpion_toxin1.
 DR Pfam; PF00537; toxin_3; 1.
 DR PRINTS; PR00284; TOXIN.
 DR PRODOM; PD000908; Scorpion_toxin1; 1.
 DR SMART; SM00505; Knoc1; 1.
 KW Neurotoxin; Sodium channel inhibitor.
 FT DISULFID 12 65 BY SIMILARITY.
 FT DISULFID 16 36 BY SIMILARITY.
 FT DISULFID 22 46 BY SIMILARITY.
 FT DISULFID 26 48 BY SIMILARITY.
 FT SEQUENCE 66 AA; 7453 MW; B5E0DB42237EF2E8 CRC64;
 Query Match 90.9%; Score 30; DB 1; Length 66;
 Best Local Similarity 80.0%; Pred. No. 11;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 22 CDSEC 26
 ID YEM9 YEAST STANDARD; PRT; 226 AA.
 AC P40064;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Hypothetical 26.2 kDa protein in GD11-COX15 intergenic region.
 GN YER139C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OC NCBI_TaxId=4932;
 RX SEQUENCE FROM N.A.
 RP STRAIN=S288c / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,
 RA Hyman R., Kayser A., Komp C., Laskari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.M.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -SIMILARITY: SOME, TO S.POMER SPAC23A1.16C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL, outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U18917; AAB64666.1; -.
 DR SGD; S0000941; YER139C.
 KM Hypothetical protein; ATP-binding.
 FT NP_BIND 74 81 ATP (POTENTIAL).
 SO SEQUENCE 226 AA; 26239 MW; E234D9ADC18C3CA CRC64;
 Query Match 90.9%; Score 30; DB 1; Length 226;
 Best Local Similarity 80.0%; Pred. No. 32;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 38 CDSEC 42


```

RESULT 4
ID Y001 CAEEL STANDARD; PRT; 331 AA.
AC P34633;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative serine/threonine-protein kinase ZK507.1 in chromosome III
   (EC 2.7.1.-).
GN ZK507.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Straden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sprat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STRONG, TO YEAST HRR25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z29116; CAA82367.2; -.
CC PIR; S40735; S40735.
CC HSSP; 006486; 1CKT.
DR WormRep; ZK507.1; CE24727.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding.
FT DOMAIN 1 270 PROTEIN_KINASE.
FT ACT SITE 116 116 BY SIMILARITY.
SQ SEQUENCE 331 AA; 38115 MW; 59615C062B84ED4 CRC64;
Query Match 90.9%; Score 30; DB 1; Length 331;
Best local Similarity 80.0%; Pred. No. 44;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
DB 44 CDAQC 48

```

```

RESULT 5
ID Y003 CAEEL STANDARD; PRT; 373 AA.
AC P34635; P34634;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 42.2 kDa protein ZK507.3 in chromosome III.
GN ZK507.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Straden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sprat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC LACKS THE ACTIVE SITE ASPARTATE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z29116; CAA82368.2; -.
CC PIR; S40737; S40737.
DR WormRep; ZK507.3; CE24728.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
KW Hypothetical protein.
SQ SEQUENCE 373 AA; 42207 MW; ECE09E332A0380EE CRC64;
Query Match 90.9%; Score 30; DB 1; Length 373;
Best local Similarity 80.0%; Pred. No. 48;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CDAEC 5
DB 86 CDAQC 90

```

```

RESULT 6
ID GUNI_HUMIN STANDARD; PRT; 402 AA.
AC P56680;

```

DT 15-JUN-1999 (Rel. 38, Created)
 DT 15-JUN-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 DE Endoglucanase I (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
 GN CEL7B.
 OS Humicola fungii.
 OC Eukaryota; Fungi; Ascomycota; microsporitic Ascomycota; Humicola.
 NCBI_TaxID=34413;
 RX MEDLINE=98437137; PubMed=9761741;
 RA Davies G.J., Dueros V., Lewis R.J., Borchert T.V., Schuelein M.;
 RT "Oligosaccharide specificity of a family 7 endoglucanase: insertion
 of potential sugar-binding subsites.";
 RL J. Biotechnol. 57:91-100(1997).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND MUTAGENESIS.
 RX MEDLINE=98437137; PubMed=9761741;
 RA Mackenzie L.F., Sulzenbacher G., Dyne C., Jones T.A., Woeldike H.F.,
 RT Schuelein M., Wilthers S.G., Davies G.J.;
 RT "Crystal structure of the family 7 endoglucanase I (cel7B) from
 Humicola insolens at 2.2 A resolution and identification of the
 RT catalytic nucleophile by trapping of the covalent glycosyl-enzyme
 intermediate.";
 RL Biochem. J. 335:409-416(1998).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCOLIBIOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC PDB; 1A39; 02-MAR-99.
 DR PDB; 2A39; 16-FEB-99.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF008840; Glyco_hydro_7; 1.
 DR ProDom: PD186135; GH_7; 1.
 KW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein;
 KM 3D-structure.
 FT MOD_RES 1 1
 FT ACT_SITE 197 197 PYRROLIDONE CARBOXYLIC ACID.
 FT ACT_SITE 202 202 NUCLEOPHILE.
 FT DISULFID 18 24 PROTON DONOR.
 FT DISULFID 51 73
 FT DISULFID 63 69
 FT DISULFID 140 365
 FT DISULFID 172 195
 FT DISULFID 176 194
 FT DISULFID 215 234
 FT DISULFID 223 228
 FT DISULFID 239 315
 FT CARBOHYD 89 89
 FT CARBOHYD 247 247 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 402 AA; 4457 MW; EDC6D31375D16535F CRC64;
 N-LINKED (GLCNAC. . .).
 Query Match 90.9%; Score 30; DB 1; Length 402;
 Best Local Similarity 80.0%; Pred. No. 51;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

AC P46237;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase type C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 GN (Cellulase) (Endoglucanase I) (EG I).
 OS Fusarium oxysporum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Fusarium.
 NCBI_TaxID=5507;
 RX MEDLINE=95047531; PubMed=7959045;
 RA Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
 RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
 RT "The use of conserved cellulase family-specific sequences to clone
 RT cellulase homologue cDNAs from Fusarium oxysporum.";
 RL Gene 150:163-167(1994).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE=97110317; PubMed=8952478;
 RA Sulzenbacher G., Drieger H., Henrichs B., Schuelein M., Davies G.J.;
 RT "Structure of the Fusarium oxysporum endoglucanase I with a
 RT nonhydrolyzable substrate analogue: substrate distortion gives rise
 RT to the preferred axial orientation for the leaving group.";
 RL Biochemistry 35:15280-15287(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=97297800; PubMed=9153437;
 RA Sulzenbacher G., Schuelein M., Davies G.J.;
 RT "Structure of the endoglucanase I from Fusarium oxysporum: native,
 RT cellobiose, and 3,4-epoxybutyl beta-D-cellobioside-inhibited forms,
 RT at 2.3-A resolution.";
 RL Biochemistry 36:5902-5911(1997).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC PDB; 1OYW; 29-OCT-97.
 DR PDB; 3OVW; 08-APR-98.
 DR PDB; 4OVW; 08-APR-98.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF008840; Glyco_hydro_7; 1.
 DR ProDom: PD186135; GH_7; 1.
 KW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; signal;
 KM 3D-structure.
 FT SIGNAL 1 18
 FT CHAIN 19 429
 FT MOD_RES 19 19
 FT ACT_SITE 215 215 PYRROLIDONE CARBOXYLIC ACID.
 FT ACT_SITE 220 220 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 36 42 PROTON DONOR (BY SIMILARITY).
 FT DISULFID 66 88
 FT DISULFID 78 84
 FT DISULFID 156 383
 FT DISULFID 190 213
 FT DISULFID 194 212
 FT DISULFID 233 252
 FT DISULFID 241 246
 FT DISULFID 257 333
 FT CARBOHYD 74 74
 FT CARBOHYD 265 265
 N-LINKED (GLCNAC. . .).
 N-LINKED (GLCNAC. . .).
 Query Match 90.9%; Score 30; DB 1; Length 402;
 Best Local Similarity 80.0%; Pred. No. 51;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

PT CARBOHYD 318 318 N-LINKED (GLCNAC. . .)
SQ SEQUENCE 429 AA; 46444 MW; FAE8A80B1F880F9 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 429;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDAEC 5
Db 190 CDAQC 194

RESULT 8
ID GUN1 HUMGR STANDARD; PRT; 435 AA.
AC 012622;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase Eg-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
EG-1.

Humicola grisea.
Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
NCBI_TaxID=5527;

SEQUENCE FROM N.A.

RC STRAIN=IFO 9854 / Var. thermoides;
RA Takahima S., Nakamura A., Hidaka M., Masaki H., Uozumi T.;

Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:

(1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
(2) EXOCELLULOHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;

(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.

-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.

-1- SUBCELLULAR LOCATION: Secreted.

-1- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
HYDROLASES).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D63516; BAA09786.1; -.

HSSP; P56680; 1A39.

InterPro; IPR001722; GH_7.

Pfam; PF00840; Glyco_hydro_7; 1.

ProDom; PD186135; GH_7; 1.

Cellulose degradation: Hydrolyase; Glycosidase; Glycoprotein; Signal.

FT CHAIN 1 20

FT MOD_RES 21 435

FT ACT_SITE 217 217

FT ACT_SITE 222 222

FT DISULFID 38 44

FT DISULFID 71 93

FT DISULFID 83 89

FT DISULFID 160 385

FT DISULFID 192 215

FT DISULFID 196 214

FT DISULFID 235 254

FT DISULFID 243 248

FT DISULFID 259 335

FT CARBOHYD 109 109

PT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
SQ SEQUENCE 435 AA; 47890 MW; 25EB3D62D96B65FA CRC64;

Query Match 90.9%; Score 30; DB 1; Length 435;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDAEC 5
Db 192 CDAQC 196

RESULT 9
ID GUN1 TRIRE STANDARD; PRT; 459 AA.
AC P07981;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase Eg-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
EG1.

Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
NCBI_TaxID=51453;

SEQUENCE FROM N.A.

RC STRAIN=VTT-D-80133;
RA MEDLINE=87106822; PubMed=2948877;
RX Penttilae M., Lehtoavaara P., Nevalainen H., Bhikhabhai R.,
RA Knowles J.K.C.;

"Homology between cellulase genes of Trichoderma reesei: complete
nucleotide sequence of the endoglucanase I gene.";

Gene 45:253-263 (1986).

SEQUENCE FROM N.A.

RC STRAIN=L27;

RA van Arsdell J.N., Kwok S., Schweickart V.L., Ladner M.B.,
RA Gelfand D.H., Innis M.A.;

"Cloning, characterization, and expression in Saccharomyces cerevisiae
of endoglucanase I from Trichoderma reesei.";

Biotechnology 5:60-64 (1987).

ACTIVE SITE GLU-149.

RA Tome P., Claysens M.;

"Identification of a functionally important carboxyl group in
cellobiohydrolase I from Trichoderma reesei.";

FEBS Lett. 243:239-243 (1989).

X-RAY CRYSTALLOGRAPHY (3.6 ANGSTROMS) OF 23-393.

MEDLINE=97467423; PubMed=9325098;

RA Kleywegt G.J., Zou J.-Y., Dym C., Davies G.J., Stanning I.,
RA Staehli J., Reinikainen T., Srisodak M., Teeri T.T., Jones T.A.;

"The crystal structure of the catalytic core domain of endoglucanase
I from Trichoderma reesei at 3.6-A resolution, and a comparison with
related enzymes.";

J. Mol. Biol. 272:383-397 (1997).

-1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:

(1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
(2) EXOCELLULOHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;

(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.

-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.

-1- SUBCELLULAR LOCATION: Secreted.

-1- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
HYDROLASES).

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M15665; AAA34212.1; -
 DR PIR; A25928; A25928.
 DR PIR; A25565; A25565.
 DR PDB; 1EG1; 19-MAR-99.
 DR InterPro; IPR001722; GH_7.
 DR InterPro; IPR001722; GH_7.
 DR Pfam; PF00734; CBM_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD166135; GH_7; 1.
 DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation; Hydrolase; glycosidase; glycoprotein; signal;
 DR 3D-structure.

FT SIGNAL 1 22
 FT CHAIN 23 459 ENDOGLUCANASE EG-1.
 FT DOMAIN 23 397 CATALYTIC.
 FT DOMAIN 398 423 LINER.
 FT DOMAIN 424 459 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 149 149
 FT ACT_SITE 218 218
 FT ACT_SITE 223 223
 FT CARBOHYD 78 78 NUCLEOPHILE.
 FT CARBOHYD 164 164 PROTON DONOR.
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 431 448 BY SIMILARITY.
 FT DISULFID 442 458 BY SIMILARITY.
 FT SEQUENCE 459 AA; 48208 MW; D235A256F808CB89 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 459;
 Best Local Similarity 80.0%; Pred. No. 57;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 Db 193 CDAQC 197

RESULT 10
 1-TRILO STANDARD; PRT; 463 AA.
 AC 012714;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase Eg-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN Egl1.
 OS Trichoderma longibrachiatum.
 OC Eukaryota; Fungi; Ascomycota; Perizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Trichoderma.
 OX NCBI_TaxID=5548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CECT 2606;
 RA Perez-Gonzalez J.A.;
 RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODIOLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLULO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X60652; CAA43059.1; -
 DR HSP; P07981; 1EG1.
 DR InterPro; IPR000254; CBD_fungal.
 DR InterPro; IPR001722; GH_7.
 DR Pfam; PF00734; CBM_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD166135; GH_7; 1.
 DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation; Hydrolase; glycosidase; glycoprotein; signal.
 DR KW
 DR SIGNAL 1 22
 DR CHAIN 23 463
 DR DOMAIN 23 397 ENDOGLUCANASE EG-1.
 DR DOMAIN 402 427 CATALYTIC.
 DR DOMAIN 428 463
 DR ACT_SITE 218 218 CELLULOSE-BINDING (BY SIMILARITY).
 DR ACT_SITE 223 223 NUCLEOPHILE (BY SIMILARITY).
 DR ACT_SITE 223 223 PROTON DONOR (BY SIMILARITY).
 DR DISULFID 435 452 BY SIMILARITY.
 DR DISULFID 446 462 BY SIMILARITY.
 DR CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR SEQUENCE 463 AA; 48337 MW; B3AC3DFD3AD2B1C CRC64;

Query Match 90.9%; Score 30; DB 1; Length 463;
 Best Local Similarity 80.0%; Pred. No. 58;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 Db 193 CDAQC 197

RESULT 11
 1-TRILO STANDARD; PRT; 471 AA.
 AC 012714;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase Eg-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN Egl1.
 OS Trichoderma longibrachiatum.
 OC Eukaryota; Fungi; Ascomycota; Perizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Trichoderma.
 OX NCBI_TaxID=5548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fritchman J.L.,
 RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Luster T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RL "The minimal genome complement of Mycoplasma genitalium.",
 Science 270:397-403 (1995).

RN [2]
 RP SEQUENCE OF 413-471 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing.";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, U39692; AAC71341.1; -;
 DR EMBL, U01796; AAD12322.1; -;
 DR TIGR, MG123; -;
 DR HYPOTHEORETICAL PROTEIN; Transmembrane; Complete proteome.
 DR TRANSMEM 10 30 POTENTIAL.
 DR SEQUENCE 471 AA; 55081 MW; C0917A8C477410CA CRC64;

 Query Match 90.9%; Score 30; DB 1; Length 471;
 Best Local Similarity 80.0%; Pred. No. 59;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 1 CDAC 5
 DB 191 CDAC 195

 RESULT 12
 ID GXUC_FUSOX STANDARD; PRT; 514 AA.
 AC P46238;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Putative exoglucanase type C precursor (EC 3.2.1.91)
 DE (Exocellulohydrolase I) (1,4-beta-cellobiohydrolase)
 DE (Beta-glucanocellobiohydrolase)
 OS Fusarium oxysporum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Mitosporic Hypocreales; Fusarium.
 OX NCBI_TaxID=5507;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=95047531; PubMed=7959045;
 RP Shepard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
 RA Hagen F.S., Uphall A., McKnight G.L., O'Hara P.J.;
 RT "The use of conserved cellulase family-specific sequences to clone
 RT cellulase homologue cDNAs from *Fusarium oxysporum*.";
 RL Gene 150:163-167(1994).
 CC -----
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, L29379; AAA65587.1; -;
 DR HSSP, P00725; 8CEL.
 DR InterPro, IPR000254; CBD_fungal.
 DR -----

DR InterPro, IPR001722; GH_7.
 DR Pfam, PF00734; CBM_1; 1.
 DR Pfam, PF00840; Glyco_hydro_7; 1.
 DR ProDom, PD001821; CBD_fungal; 1.
 DR ProDom, PD186135; GH_7; 1.
 DR SMART, SM00236; ICBF; 1.
 DR PROSITE, PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; signal.
 KW CELLULOSE DEGRADATION; HYDROLASE; 1.
 FT SIGNAL 1 17
 FT CHAIN 18 514
 FT DOMAIN 18 439
 FT DOMAIN 440 482
 FT DOMAIN 483 514
 FT ACT SITE 229 229
 FT ACT SITE 234 234
 FT DISULFID 486 503
 FT DISULFID 497 513
 FT CARBOHYD 287 287
 FT CARBOHYD 490 490
 FT SEQUENCE 514 AA; 54704 MW; 6A4617323A46E062 CRC64;

 Query Match 90.9%; Score 30; DB 1; Length 514;
 Best Local Similarity 80.0%; Pred. No. 63;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 1 CDAC 5
 DB 189 CDAC 193

 RESULT 13
 ID GXU1_NEUCR STANDARD; PRT; 516 AA.
 AC P38676;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Exoglucanase 1 precursor (EC 3.2.1.91) (Exocellulohydrolase 1)
 DE (1,4-beta-cellobiohydrolase)
 GN CBH-1.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=74-OR23-1A;
 RP MEDLINE=95369725; PubMed=7642129;
 RA Taleb F., Radford A.;
 RT "The cellulase complex of *Neurospora crassa*: cbh-1 cloning,
 RT sequencing and homologues.";
 RL Gene 161:137-138(1995).
 CC -----
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X77778; CAA54815.1; -;
 DR PIR, S42093; S42093.
 DR HSSP, P00725; 8CEL.
 DR InterPro, IPR000254; CBD_fungal.
 DR InterPro, IPR001722; GH_7.
 DR -----

```

DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; GH_7; 1.
DR SMART; SM00236; fCBD; 1.
KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 17
FT CHAIN 18 516
FT DOMAIN 18 445
FT DOMAIN 18 445
FT DOMAIN 446 480
FT ACT_SITE 481 516
FT ACT_SITE 223 223
FT ACT_SITE 228 228
FT DISULFID 488 505
FT DISULFID 499 515
SQ SEQUENCE 516 AA; 54471 MW; 38559406EA81900 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 516;
st Local Similarity 80.0%; Pred. No. 63;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CDAEC 5
Db 184 CDAQC 188

RESULT 14
GUXL HUMGR STANDARD; PRT; 525 AA.
AC P15828;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-
DE beta-cellobiohydrolase) (beta-glucanocellobiohydrolase).
GN CBH-1.
OS Humicola grisea.
OC Eukaryota; Fungi; Ascomycota; microsporid Ascomycota; Humicola.
OX NCBI_Taxid=5527;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Var. thermoidae;
RA MEDLINE=90175006; PubMed=2308855;
RA de Oliveira Alzevedo M., Radford A.;
RT "Sequence of chb-1 gene of Humicola grisea var. thermoidae.";
RT Nucleic Acids Res. 18:668-668(1990).
RL
-1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
GENERALY REQUIRES THREE TYPES OF HYDROLITIC ENZYMES:
(1) ENDOLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
(2) EXOCYLOBIHYDROLASES THAT CUT THE DISACCHARIDE POLYMER CHAIN;
(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
-1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
in cellulose and cellooligosaccharides, releasing cellobiose from the non-
reducing ends of the chains.
-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
HYDROLASES).
-1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; X17258; CAA35159.1; -.
CC PIR; S08240; S08240.
CC HSSP; P00725; BCEL.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001722; GH_7.

```

```

DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; GH_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 525
FT DOMAIN 19 467
FT DOMAIN 468 489
FT DOMAIN 490 525
FT ACT_SITE 231 231
FT ACT_SITE 236 236
FT CARBOHYD 289 289
FT DISULFID 497 514
FT DISULFID 508 524
SQ SEQUENCE 525 AA; 55693 MW; A6684D9CF81E090 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 525;
st Local Similarity 80.0%; Pred. No. 64;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CDAEC 5
Db 191 CDAQC 195

RESULT 15
PEGA HUMAN STANDARD; PRT; 528 AA.
ID PEGA_HUMAN
AC Q96134;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Protein phosphatase 1 regulatory inhibitor subunit 16A.
GN PPR16A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL
-1- FUNCTION: Inhibits protein phosphatase 1 activity toward
phosphorylase, myosin light chain and myosin substrates (By
similarity).
-1- SUBUNIT: Binds ppl.
-1- SUBCELLULAR LOCATION: Attached to the plasma membrane (By
similarity).
-1- CAUTION: Ref.1 sequence may be incorrect due to a cloning
artifact.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AK090471; BAC03452.1; ALT_SEQ.
DR EMBL; BC007854; AA07854.1; -.

```

DR EMBL; BC015896; AAH15896.1; -.
 DR Genew; HGNC:14941; PEP1R16A.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001230; Prem1_site.
 DR Pfam; PF000023; ank; 5.
 DR PROSITE; PS00294; PRENYLATION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR ANK repeat; Repeat; Coiled coil; Lipoprotein; Prenylation; Membrane.
 FT DOMAIN 18 45 COILED COIL (POTENTIAL).
 FT REPEAT 70 99 ANK 1.
 FT REPEAT 103 132 ANK 2.
 FT REPEAT 136 165 ANK 3.
 FT REPEAT 231 260 ANK 4.
 FT REPEAT 264 293 ANK 5.
 FT LIPID 525 525 FARNESYL (POTENTIAL).
 FT CONFLICT 34 71 MISSING (IN REF. 1).
 FT CONFLICT 249 249 A -> P (IN REF. 1).
 FT CONFLICT 272 272 A -> P (IN REF. 1).
 SO SEQUENCE 528 AA; 57811 MW; A5A2D5E1A16A438C CRC64;
 Key Match 90.9%; Score 30; DB 1; Length 528;
 Best Local Similarity 80.0%; Pred. No. 64;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 Db 133 CDSEC 137
 RESULT 16
 L112 CAEBL STANDARD; PRT; 1429 AA.
 ID L112 CAEBL
 AC P14585;
 DT 01-JUN-1990 (Rel. 13, Created)
 DT 01-JUN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lin-12 protein precursor.
 GN LIN-12 OR R107.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RX MEDLINE=88334747; PubMed=3419531;
 RA "Yochem J., Weston K., Greenwald I.;
 "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
 protein with overall similarity to Drosophila Notch.";
 Nature 335:547-550(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favell A., Frazer A.,
 Fulton L., Garner A., Green P., Hawkins T., Hillier L., Jier M.,
 Johnston L., Jones M., Kershaw J., Kizsten J., Laister N.,
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showlken R.,
 Sims M., Smaldon N., Smith A., Smith M., Sonhammer B., Staden R.,
 Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 Waterston R., Watson A., Weinstock L., Wilkinson-Sprat J.,
 Woilman P.;
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 Nature 368:32-38(1994).
 CC -1- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT
 REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12
 ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES
 EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.
 CC -1- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M12069; AAA70191.1; -.
 DR EMBL; Z14092; CAA78474.1; -.
 DR PIR; S06434; S06434.
 DR HSSP; P00740; IEDM.
 DR WormBep; R107.8; CE00274.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR00742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00008; EGF; 13.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 4.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00001; EGF_like; 10.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 12.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS01187; EGF_CA; 2.
 KM Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 1429
 FT DOMAIN 16 908
 FT TRANSMEM 909 931
 FT DOMAIN 932 1429
 FT DOMAIN 20 61
 FT DOMAIN 114 150
 FT DOMAIN 152 246
 FT DOMAIN 201 285
 FT DOMAIN 250 323
 FT DOMAIN 287 323
 FT DOMAIN 323 363
 FT DOMAIN 365 402
 FT DOMAIN 404 441
 FT DOMAIN 449 492
 FT DOMAIN 503 541
 FT DOMAIN 543 579
 FT DOMAIN 582 619
 FT REPEAT 635 669
 FT REPEAT 670 710
 FT REPEAT 711 750
 FT REPEAT 1093 1122
 FT REPEAT 1126 1158
 FT REPEAT 1162 1194
 FT REPEAT 1206 1236
 FT REPEAT 1240 1269
 FT DISULFID 24 35
 FT DISULFID 29 49
 FT DISULFID 51 60
 FT DISULFID 118 129
 FT DISULFID 123 138
 FT DISULFID 140 149
 POTENTIAL.
 LIN-12 PROTEIN.
 EXTRACELLULAR (POTENTIAL).
 CYTOPLASMIC (POTENTIAL).
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5.
 EGF-LIKE 6.
 EGF-LIKE 7.
 EGF-LIKE 8.
 EGF-LIKE 9.
 EGF-LIKE 10.
 EGF-LIKE 11.
 EGF-LIKE 12.
 EGF-LIKE 13.
 LIN/NOTCH 1.
 LIN/NOTCH 2.
 LIN/NOTCH 3.
 ANK 1.
 ANK 2.
 ANK 3.
 ANK 4.
 ANK 5.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

FT DISULFID 156 169 BY SIMILARITY.
 FT DISULFID 163 178 BY SIMILARITY.
 FT DISULFID 180 189 BY SIMILARITY.
 FT DISULFID 205 227 BY SIMILARITY.
 FT DISULFID 221 234 BY SIMILARITY.
 FT DISULFID 236 245 BY SIMILARITY.
 FT DISULFID 254 264 BY SIMILARITY.
 FT DISULFID 259 273 BY SIMILARITY.
 FT DISULFID 275 284 BY SIMILARITY.
 FT DISULFID 291 302 BY SIMILARITY.
 FT DISULFID 296 311 BY SIMILARITY.
 FT DISULFID 313 322 BY SIMILARITY.
 FT DISULFID 327 339 BY SIMILARITY.
 FT DISULFID 351 351 BY SIMILARITY.
 FT DISULFID 353 362 BY SIMILARITY.
 FT DISULFID 369 381 BY SIMILARITY.
 FT DISULFID 375 390 BY SIMILARITY.
 FT DISULFID 392 401 BY SIMILARITY.
 FT DISULFID 408 419 BY SIMILARITY.
 FT DISULFID 413 429 BY SIMILARITY.
 FT DISULFID 431 440 BY SIMILARITY.
 FT DISULFID 507 518 BY SIMILARITY.
 FT DISULFID 512 529 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 547 558 BY SIMILARITY.
 FT DISULFID 552 567 BY SIMILARITY.
 FT DISULFID 569 578 BY SIMILARITY.
 FT DISULFID 586 597 BY SIMILARITY.
 FT DISULFID 591 607 BY SIMILARITY.
 FT DISULFID 609 618 BY SIMILARITY.
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1429 AA; 157115 MW; 255EDD7A62C025DB CRC64;

Query Match 90.9%; Score 30; DB 1; Length 1429;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDABC 5
 652 CDADC 656

RESULT 17
 SIXS ANDAU STANDARD; PRT; 61 AA.
 AC P81564;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Insect toxin 5 (Insect toxin Anaitis).
 OS Androctonus australis hector (Sahara scorpion).
 OC Eukaryote; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
 OC Buttholde; Buttholde; Androctonus.
 NCBI_Taxid=70175;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RX MEDLINE=97352548; PubMed=9208943;
 RA Nakagawa Y., Lee Y.M., Lehmberg E., Herrmann R.,
 RT Moskowitz H., Jones A.D., Hammock B.D.;
 RT "Anti-Insect toxin 5 (AaIT5) from Androctonus australis";
 RL Eur. J. Biochem. 246:496-501(1997).
 CC -!- FUNCTION: BINDS TO SODIUM CHANNELS AND INHIBITS THE INACTIVATION
 OF THE ACTIVATED CHANNELS, THEREBY BLOCKING NEURONAL TRANSMISSION.
 CC THIS TOXIN IS ACTIVE ONLY ON INSECTS AND SHOWS A DISTINCT INSECT

CC SPECIFICITY. IT IS VERY POTENT AGAINST THE TOBACCO BUDMORM,
 CC HELIOTHIS VIRESCENS (100% LETHAL DOSE < 1.8 MICROG/100 MG BODY
 CC MASS).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.
 CC ALPHA-TOXIN SUBFAMILY.
 CC HSSP; P01494; 25N3.
 DR InterPro; IPR003614; Knott.
 DR InterPro; IPR002061; Scorpion_toxinL.
 DR Pfam; PF00537; toxin_3; 1.
 DR ProDom; PD000908; Scorpion_toxinL; 1.
 DR SMART; SM00505; Knott1; 1.
 DR Neurotoxin; Sodium channel inhibitor.
 KW DISULFID 10 60 BY SIMILARITY.
 FT DISULFID 14 35 BY SIMILARITY.
 FT DISULFID 21 42 BY SIMILARITY.
 FT DISULFID 25 44 BY SIMILARITY.
 SQ SEQUENCE 61 AA; 6889 MW; CCF5E269AFE346AA CRC64;

Query Match 87.9%; Score 29; DB 1; Length 61;
 Best Local Similarity 80.0%; Pred. No. 16;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CDABC 5
 21 CDTEC 25

RESULT 18
 ITX1_TEGAG STANDARD; PRT; 68 AA.
 AC O46166;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Paralytic insecticidal toxin 1 precursor (Taityx-1).
 GN ITX1.
 OS Tegenaria agrestis (Spider).
 OC Eukaryote; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Encelegrinae; Agelenidae; Tegenaria.
 NCBI_Taxid=73900;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Venom gland;
 RX MEDLINE=98251507; PubMed=9589602;
 RA Johnson J.H., Bloomquist J.R., Krupcho K.J., Kral R.M. Jr.,
 RA Trovato R., Eppler K.G., Morgan T.K., Delmar E.G.;
 RT "Novel insecticidal peptides from Tegenaria agrestis spider venom may
 RT have a direct effect on the insect central nervous system";
 RL Arch. Insect Biochem. Physiol. 38:19-31(1998).
 CC -!- FUNCTION: TOXIN THAT PARALYZES INSECTS. MAY HAVE A DIRECT EFFECT
 ON THE INSECT CENTRAL NERVOUS SYSTEM.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; AJ224127; CA11839.1; -.
 KW Toxin; Amidation; Signal.
 FT SIGINL 1 17 POTENTIAL.
 FT CHAIN 18 67 PARALYTIC INSECTICIDAL TOXIN 1.
 FT MOD RES 67 67 AMIDATION (G-68 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 68 AA; 7741 MW; 19578D59D92F268 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 68;
 Best Local Similarity 80.0%; Pred. No. 18;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|:|:|
DB 49 CEAC 53

RESULT 19
ITX2_TEGAG STANDARD; PRT; 68 AA.
AC 046167;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Paralytic insecticidal toxin 2 precursor (Taixt-2).
GN ITX2.
OS Tegenaria agrestis (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Agelenidae; Tegenaria.
ON NCBI_TaxID=73900;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Venom gland;
MEDLINE=98251507; PubMed=9589602;
RA Johnson J.H., Bloomquist J.R., Kratoch K.J., Kral R.M. Jr.,
RA Trovato R., Eppler K.G., Morgan T.K., Delmar E.G.;
RT "Novel insecticidal peptides from Tegenaria agrestis spider venom may
have a direct effect on the insect central nervous system.";
RL Arch. Insect Biochem. Physiol. 38:19-31(1998).
CC -1- FUNCTION: TOXIN THAT PARALYZES INSECTS. MAY HAVE A DIRECT EFFECT
ON THE INSECT CENTRAL NERVOUS SYSTEM.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ224128; CA11840.1; --
KW Toxin; Amidation; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 67 PARALYTIC INSECTICIDAL TOXIN 2.
FT MOD_RES 67 67 AMIDATION (G-68 PROVIDE AMIDE GROUP).
SQ SEQUENCE 68 AA; 7762 MW; 8EC9F344FEFD126D CRC64;

Query Match 87.9%; Score 29; DB 1; Length 68;
Best Local Similarity 80.0%; Pred. No. 18;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|:|:|
DB 49 CEAC 53

RESULT 20
ITX3_TEGAG STANDARD; PRT; 68 AA.
AC 046168;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Paralytic insecticidal toxin 3 precursor (Taixt-3).
GN ITX3.
OS Tegenaria agrestis (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Agelenidae; Tegenaria.
ON NCBI_TaxID=73900;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Venom gland;

RX MEDLINE=98251507; PubMed=9589602;
RA Johnson J.H., Bloomquist J.R., Kratoch K.J., Kral R.M. Jr.,
RA Trovato R., Eppler K.G., Morgan T.K., Delmar E.G.;
RT "Novel insecticidal peptides from Tegenaria agrestis spider venom may
have a direct effect on the insect central nervous system.";
RL Arch. Insect Biochem. Physiol. 38:19-31(1998).
CC -1- FUNCTION: TOXIN THAT PARALYZES INSECTS. MAY HAVE A DIRECT EFFECT
ON THE INSECT CENTRAL NERVOUS SYSTEM.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ224129; CA11841.1; --
KW Toxin; Amidation; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 67 PARALYTIC INSECTICIDAL TOXIN 3.
FT MOD_RES 67 67 AMIDATION (G-68 PROVIDE AMIDE GROUP).
SQ SEQUENCE 68 AA; 7705 MW; 8EC9F34D664126D CRC64;

Query Match 87.9%; Score 29; DB 1; Length 68;
Best Local Similarity 80.0%; Pred. No. 18;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
|:|:|
DB 49 CEAC 53

RESULT 21
SC11 MESMA STANDARD; PRT; 85 AA.
AC 09NJC7; 09NJP8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alpha-neurotoxin TX11 precursor (Toxin BmtX11) (Alpha toxin 2).
OS Mesobuthus martenisi (Menchurian scorpion) (Buthus martenisi).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC Buthoidea; Buthidae; Mesobuthus.
ON NCBI_TaxID=34649;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Venom gland;
RC MEDLINE=20317244; PubMed=10858508;
RX Zhu S.-Y., Li W.-X., Zeng X.-C., Liu H., Jiang D.-H., Mao X.;
RA "Nine novel precursors of Buthus martenisi scorpion alpha-toxin
homologues.";
RL Toxicon 38:1653-1661(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Ye J., Chen J., Zuo X., Ji Y.;
RT "Cloning and characterization of cDNA sequences encoding two novel
alpha toxin precursors from the Chinese scorpion Buthus martenisi
Karsch.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBD databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155364; AAF31478.1; --
DR EMBL; AF132974; AAF31295.1; --

DR EMBL; AF288608; AAC00581.1; -.
 DR HSSP; P01484; IAH0.
 DR InterPro; IPR003614; Knto1.
 DR InterPro; IPR001219; Neurotoxin.
 DR InterPro; IPR002061; Scorpion_toxinL.
 DR Pfam; PF00537; toxin_3; 1.
 DR PRINTS; PR00284; TOXIN.
 DR ProDom; PD000908; Scorpion_toxinL; 1.
 DR SMART; SM00505; Knto1; 1.
 KM Neurotoxin; Sodium channel inhibitor; Signal.
 FT SIGNAL 19
 FT CHAIN 1
 FT DISULFID 31 85 ALPHA-NEUROTOXIN TX11.
 FT DISULFID 35 55 BY SIMILARITY.
 FT DISULFID 41 65 BY SIMILARITY.
 FT DISULFID 45 67 BY SIMILARITY.
 FT VARIANT 4 7 MVII -> LVFF (IN ISOFORM TX11P).
 FT VARIANT 13 13 V -> L (IN ISOFORM TX11P).
 FT SEQUENCE 85 AA; 9524 MW; 565390BBE71806E CRC64;
 Query Match 87.9%; Score 29; DB 1; Length 85;
 Best Local Similarity 80.0%; Pred. No. 21;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAC 5
 DB 41 CDGEC 45
 RESULT 22
 ID SCX1_CENSC STANDARD; PRT; 87 AA.
 AC P01492; Q95WB7; Q95WB8; Q95WC6; Q95WC7;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurotoxin 1 precursor (CseV1) (CSE V1).
 OS Centruroides sculpturatus (Bark scorpion).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
 OC Butthoidea; Butthidae; Centruroides.
 OC NCBI_TaxId=6879;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM CSEV1B; CSEV1C; CSEV1D AND CSEV1E).
 RC TISSUE=Venom gland;
 RX MEDLINE=21486965; PubMed=11600153;
 RA Corona M., Valdez-Cruz N.A., Merino E., Zurita M., Possani L.D.;
 RT "Genes and peptides from the scorpion Centruroides sculpturatus Ewing,
 that recognize Na(+) -channels";
 RT Toxicon 39:1893-1898(2001).
 RN [2]
 RP SEQUENCE OF 20-84 (ISOFORM CSEV1B).
 RC TISSUE=Venom;
 RX MEDLINE=75163395; PubMed=4460885;
 RA Babin D.R., Watt D.D., Goos S.M., Mlejnek R.V.;
 RT "Amino acid sequences of neurotoxic protein variants from the venom
 of Centruroides sculpturatus Ewing.";
 RT Arch. Biochem. Biophys. 164:694-706(1974).
 RN [3]
 RP REVISIONS TO 44-46 AND 83-84, AND STRUCTURE BY NMR.
 RC TISSUE=Venom;
 RX MEDLINE=94162260; PubMed=8117707;
 RA Lee W., Jablonaky M.J., Watt D.D., Krishna N.R.;
 RT "Proton nuclear magnetic resonance and distance geometry/simulated
 annealing studies on the variant-1 neurotoxin from the New World
 scorpion Centruroides sculpturatus Ewing.";
 RT Biochemistry 33:2468-2475(1994).
 RN [4]
 RP FUNCTION: Binds to sodium channels and inhibits the inactivation
 of the activated channels, thereby blocking neuronal transmission.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.
 CC BETA-TOXIN SUBFAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF38454; AAL23422.1; -.
 DR EMBL; AF38455; AAL23423.1; -.
 DR EMBL; AF38463; AAL23431.1; -.
 DR EMBL; AF38464; AAL23432.1; -.
 DR FTR; A01752; NTSRIC.
 DR PDB; 1VNA; 31-JAN-94.
 DR PDB; 1VNB; 31-JAN-94.
 DR InterPro; IPR003614; Knto1.
 DR InterPro; IPR001219; Neurotoxin.
 DR InterPro; IPR002061; Scorpion_toxinL.
 DR Pfam; PF00537; toxin_3; 1.
 DR PRINTS; PR00284; TOXIN.
 DR ProDom; PD000908; Scorpion_toxinL; 1.
 DR SMART; SM00505; Knto1; 1.
 KM Toxin; Neurotoxin; Sodium channel inhibitor; Signal; Amidation;
 KM 3D-structure. 1 19
 FT SIGNAL 1 19
 FT PEPTIDE 20 84
 FT MOD RES 84 84
 FT DISULFID 31 84
 FT DISULFID 35 60
 FT DISULFID 44 65
 FT DISULFID 48 67
 FT VARIANT 6 6
 FT VARIANT 9 14
 FT VARIANT 26 26
 FT VARIANT 29 29
 FT VARIANT 40 40
 FT VARIANT 43 43
 FT VARIANT 46 46
 FT VARIANT 46 46
 FT VARIANT 45 45
 FT CONFLICT 21 23
 FT STRAND 21 25
 FT STRAND 27 29
 FT TURN 27 29
 FT STRAND 32 32
 FT STRAND 34 34
 FT STRAND 39 39
 FT STRAND 44 49
 FT HELIX 44 49
 FT TURN 51 53
 FT STRAND 57 60
 FT STRAND 64 68
 FT STRAND 79 79
 FT TURN 80 81
 FT STRAND 82 82
 SQ SEQUENCE 87 AA; 9636 MW; 9237B1FCBBF0E658 CRC64;
 Query Match 87.9%; Score 29; DB 1; Length 87;
 Best Local Similarity 80.0%; Pred. No. 22;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAC 5
 DB 44 CDTEC 48
 RESULT 23
 ID Y609_METTH STANDARD; PRT; 147 AA.
 AC P72011;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein MTH609.
 GN MTH609.

OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 RX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98080610; PubMed=9419225;
 RA Koga Y., Kyuragi T., Nishihara M., Sone N.;
 RT "Old archael and bacterial cells arise independently from noncellular
 precursors? A hypothesis stating that the advent of membrane
 phospholipid with enantiomeric glycerophosphate backbones caused the
 separation of the two lines of descent.";
 RT J. Mol. Evol. 46:54-63(1998).
 RL [2]
 RP ERRATUM.
 RX PubMed=9797414;
 RA Koga Y., Kyuragi T., Nishihara M., Sone N.;
 RL J. Mol. Evol. 47:631-631(1998).
 RN [3]
 CC SEQUENCE FROM N.A.
 CC STRAIN=Delta H;
 CC MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doncette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Baehrizen R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
 RA Spadafora R., Vcare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Carnuo A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delta: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- SIMILARITY: BELONGS TO THE UPF0179 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; D88555; BAAL3645.1; ALT INIT.
 DR EMBL; AEO00842; AAB85115.1; -
 DR InterPro: IPR005369; UPF0179.
 DR Pfam: PF03684; UPF0179.1.
 DR KMW Hypothetical protein; Complete proteome.
 DR HYPOTHETICAL 147 AA; 16773 MW; 7583D432C918E028 CRC64;
 SQ
 CC
 CC Query Match 87.9%; Score 29; DB 1; Length 147;
 CC Best Local Similarity 80.0%; Pred. No. 34;
 CC Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 96 CDVEC 100
 RESULT 24
 BMEP_RAT
 ID BMEP_RAT STANDARD; PRT; 207 AA.
 AC Q04906;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Bone morphogenetic protein 6 precursor (BMP-6) (VG-1-related protein)
 DE (VGR-1) (Fragment).
 GN BMP6 OR BMP-6 OR VGR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Lewis; TISSUE=Brain;
 RX MEDLINE=93085758; PubMed=1453478;
 RA Suetsmann U., Meyermann R., Schluesener H.J.;
 RT "Cloning of a novel TGF-beta related cytokine, the vgr. from rat
 RT brain: cloning of and comparison to homologous human cytokines.";
 RL J. Neurosci. Res. 33:142-147(1992).
 CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; X58830; CAA1634.1; -
 DR PIR; S37618; S37618.
 DR HSSP; P18075; IBMP.
 DR InterPro: IPR001839; TGF-beta.
 DR InterPro: IPR001111; TGF-beta.
 DR Pfam; PF000688; TGF-beta; 1.
 DR Pfam; PF00688; TGF-beta; 1.
 DR ProDom; PD000357; TGF-beta; 1.
 DR SMART; SM00204; TGF-beta; 1.
 DR PROSITE; PS00250; TGF-beta; 1.
 KW Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
 FT NON TER 1
 FT PROPEP 1
 FT CHAIN 76
 FT DISULFID 106
 FT DISULFID 135
 FT DISULFID 139
 FT DISULFID 171
 FT CARBOHYD 80
 FT CARBOHYD 98
 FT CARBOHYD 148
 SQ SEQUENCE 207 AA; 23225 MW; D5D83C3898213F51 CRC64;
 CC
 CC Query Match 87.9%; Score 29; DB 1; Length 207;
 CC Best Local Similarity 80.0%; Pred. No. 46;
 CC Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 135 CDSEC 139
 RESULT 25
 FBK2_ARATH
 ID FBK2_ARATH STANDARD; PRT; 223 AA.
 AC Q22870; Q940C0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable FRP-type peptidyl-prolyl cis-trans isomerase 2, chloroplast
 DE precursor (EC 5.2.1.8) (PPIase) (Notase).
 GN AT2G43560 OR TIO24.30.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Maon T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,

RA Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Niernm W.C., White O., Eissen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
RL *thaliana*.";
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M.U., Yamamura Y., Yu G.,
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA Nguyen M., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Koyama E., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A.M., Tracy S.E., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RP "Full length cDNA of gene T1024.30/At2g43560 (GI:2289010).";
RN Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 77-96.
RX MEDLINE=21864138; PubMed=11719511;
RA Schubert M., Petersson U.A., Haas B.J., Funk C., Schroeder W.P.,
RA Kieselbach T.;
RT "Proteome map of the chloroplast lumen of *Arabidopsis thaliana*.";
RL J. Biol. Chem. 277:8354-8365 (2002).
CC -1- FUNCTION: Phases accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides (by similarity).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AC002335; AAB64339.2; -;
CC EMBL; AY056125; AAL07204.1; -;
CC HSSP; P20081; 1YAT.
CC InterPro; IPR001179; FKBP_PPIase.
CC Pfam; PF00254; FKBP; 1.
DR PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR PROSITE; PS00454; FKBP_PPIASE_2; FALSE_NEG.
DR PROSITE; PS00059; FKBP_PPIASE_3; 1.
KM Isomerase; Rotomase; Chloroplast; Thylakoid; Transit peptide.
FT TRANSIT 1 36 CHLOROPLAST (POTENTIAL).
FT TRANSIT 37 77 THYLAKOID.
FT CHAIN 78 223 PROBABLE FKBP-TYPE PEPTIDYL-PROLYL CIS-
FT TRANSIT 78 223 TRANS ISOMERASE 2.
FT DOMAIN 124 216 PPIASE, FKBP-TYPE.
SQ SEQUENCE 223 AA; 23564 MM; BR3A48D0CIA59C8 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 223;
Best Local Similarity 80.0%; Pred. No. 48;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 88 CEAC 92

RESULT 26
RB33_CAEEL STANDARD; PRT; 307 AA.
AC Q20365; Q94154; RN

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Putative RAS-related protein F43D9.2.
GN F43D9.2.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; *Caenorhabditis*.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briceol N2;
RA Wilkinson U.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 154-219 FROM N.A.
RA Nonet M.L., Stanton J.E., Kilgard M.P., Fergestad T.,
RA Jorgensen E., Hartweig E., Horvitz H.R., Meyer B.J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.
CC MOST SIMILAR TO MAMMALIAN RAB33A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z35640; CAA84705.1; -;
CC EMBL; U68264; AAB16979.1; -;
CC WormRep; F43D9.2; CE00992.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_tnsfmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTNSFRMG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small GTP; 1.
KM Hypothetical protein; GTP-binding; Lipoprotein; Prenylation.
FT NE_BIND 107 114 GTP (BY SIMILARITY).
FT NE_BIND 155 159 GTP (BY SIMILARITY).
FT NE_BIND 217 220 GTP (BY SIMILARITY).
FT DOMAIN 129 137 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 306 306 GERANYL-GERANYL (POTENTIAL).
FT LIPID 307 307 GERANYL-GERANYL (POTENTIAL).
SQ SEQUENCE 307 AA; 34344 MM; IB23B78F7DADF0 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 307;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 219 CDVEC 223

RESULT 27
HEMA_IHAR STANDARD; PRT; 344 AA.
AC P28730;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain]
DE (Fragment).
GN HA.
OS Influenza A virus (strain A/Hardin/1/88).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=31661;
RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=9216746; PubMed=1538194;
RA Guo Y., Xu X., Cox N.J.;
RT "Human influenza A (H1N2) viruses isolated from China.";
RL J. Gen. Virol. 73:383-388(1992).
CC CC
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC (HAI AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L19006; AAA43175.1; -.
PIR; P00319; P00319.
InterPro; IPR001364; Hemagglutn.
Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 344 344
SQ SEQUENCE 344 AA; 38563 MW; 4BAAB1B343087B89 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 344;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 292 CDAKC 296

RESULT 28
HAI_HAI
HAI_HAI STANDARD; PRT; 344 AA.
AC P12550;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain]
DE (Fragment).
GN HA.
OS Influenza A virus (strain A/Taiwan/1/86).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_Taxid=11465;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87197215; PubMed=3572359;
RA Bootman J.S., Daniels R., Robertson J.S.;
RT "Sequence analysis of the haemagglutinin of A/Taiwan/1/86, a new
RT variant of human influenza A(H1N1) virus.";
RL J. Gen. Virol. 68:1205-1208(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HAI AND HA2) LINKED BY A DISULFIDE BOND.

```

```

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17224; CAA35097.1; -.
PIR; A26765; HAIYTA.
InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; signal.
FT SIGNAL 1 17
FT CHAIN 18 343 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 344 344
SQ SEQUENCE 344 AA; 38434 MW; 189C4F3D55D1EC68 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 344;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 292 CDAKC 296

RESULT 29
HAI_HAI
HAI_HAI STANDARD; PRT; 344 AA.
AC P28731;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain]
DE (Fragment).
GN HA.
OS Influenza A virus (strain A/Xianfeng/3/89).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_Taxid=31665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9216746; PubMed=1538194;
RA Guo Y., Xu X., Cox N.J.;
RT "Human influenza A (H1N2) viruses isolated from China.";
RL J. Gen. Virol. 73:383-388(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HAI AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; L19005; AAA9877.1; -.
DR PIR; P00321; P00321.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 344 344
FT SEQUENCE 344 AA; 38568 MW; BD9EG3EAD1FE66B CRC64;

Query Match 87.9%; Score 29; DB 1; Length 344;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDABC 5
Db 292 CDABC 296

RESULT 30
ID DNAB CAUCR STANDARD; PRT; 385 AA.
AC P22305;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chapterone protein dnab.
GN DNAB OR CC0011.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
OC [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
Koltonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
Unterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
"Complete genome sequence of Caulobacter crescentus.",
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
[2]
RP SEQUENCE OF 1-193 FROM N.A.
RX MEDLINE=90264295; PubMed=2345134;
RA Gomes S.L., Guber J.W., Shapiro L.;
RT "Expression of the Caulobacter heat shock gene dnab is
RT developmentally controlled during growth at normal temperatures.";
RL J. Bacteriol. 172:3051-3059(1990).
CC -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC THE ATPASE ACTIVITY OF DNAB (BY SIMILARITY).
CC -1- COPOLYMER BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DNAB FAMILY.
CC -1- SIMILARITY: CONTAINS 1 Z DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005676; AAC2199.1; -.
DR EMBL; M55224; AAA62724.1; -.
DR PIR; B35388; B35388.
DR HSSP; P08622; 1BQZ.
DR TIGR; CC0011; -.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXXCXXG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF00684; DnaJ_CXXCXXG; 1.
DR Pfam; PF01556; DnaJ_C; 1.
DR PRINTS; PR00625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS00076; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXXCXXG; 1.
KW Chapterone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KW Complete proteome.
FT DOMAIN 1 70 J-DOMAIN.
FT DOMAIN 77 118 GLY-RICH.
FT REPEAT 148 155 CXXCXXG MOTIF.
FT REPEAT 165 172 CXXCXXG MOTIF.
FT REPEAT 187 194 CXXCXXG MOTIF.
FT REPEAT 201 208 CXXCXXG MOTIF.
FT METAL 148 148 ZINC 1 (BY SIMILARITY).
FT METAL 151 151 ZINC 2 (BY SIMILARITY).
FT METAL 165 165 ZINC 2 (BY SIMILARITY).
FT METAL 168 168 ZINC 2 (BY SIMILARITY).
FT METAL 187 187 ZINC 2 (BY SIMILARITY).
FT METAL 190 190 ZINC 2 (BY SIMILARITY).
FT METAL 201 201 ZINC 1 (BY SIMILARITY).
FT METAL 204 204 ZINC 1 (BY SIMILARITY).
FT CONFLICT 22 23 P -> S (IN REF. 2).
FT CONFLICT 58 58 P -> S (IN REF. 2).
FT CONFLICT 77 77 GUNGGPG -> AGQRGRN (IN REF. 2).
FT CONFLICT 110 114 ROSNA -> AVOR (IN REF. 2).
FT CONFLICT 142 147 VPAAMT -> IPRHEP (IN REF. 2).
FT CONFLICT 161 166 SPVCG -> OPLCLR (IN REF. 2).
FT CONFLICT 173 173 R -> P (IN REF. 2).
FT CONFLICT 185 188 RGCP -> AA (IN REF. 2).
SQ SEQUENCE 385 AA; 40965 MW; 93018119462B2C5 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 385;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5
Db 295 CDABC 299

RESULT 31
ID UNIV_STRPU STANDARD; PRT; 395 AA.
AC P48970;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Univin precursor.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoda; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
OC [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95046897; PubMed=7958442;

```

RA Stenzel P., Angerer L.M., Smith B.J., Angerer R.C., Vale W.W.;
 RT "The univin gene encodes a member of the transforming growth
 factor-beta superfamily with restricted expression in the sea urchin
 embryo.";
 RL Dev. Biol. 166:149-158(1994).
 RN [2]
 RP REVISIONS TO N-TERMINUS.
 RA Angerer L.M., Stenzel P.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD HAVE A CRITICAL ROLE IN EARLY DEVELOPMENTAL
 DECISIONS IN THE SEA URCHIN EMBRYO.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: HIGHEST LEVELS OF EXPRESSION IN THE EGG AND
 PREHATCHING BLASTULA. DURING LATE CLEAVAGE STAGES, IT ACCUMULATES
 PROGRESSIVELY TO A CIRCUMQUATORIAL BAND. DURING GASTRULATION IT
 IS DETECTED PRIMARILY IN THE PRESUMPTIVE FOREGUT AND CLINTED
 BAND. BY PLUTERUS STAGE, IT IS DETECTED ONLY IN THESE CELL TYPES.
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, U0533; AAA57553.2; -.
 DR HSSP; P12643; 3BMP.
 DR InterPro; IPR002405; Inhibin_alpha.
 DR InterPro; IPR001839; TGFb.
 DR InterPro; IPR001111; TGFb_N.
 DR Pfam; PF000019; TGF-beta; 1.
 DR Pfam; PF00688; TGFb propeptide; 1.
 DR PRINTS; PRO0669; INHIBIN.
 DR ProDom; PD000357; TGFb; 1.
 DR SMART; SM00204; TGFb; 1.
 DR PROSITE; PS00250; TGF_BETA_1; 1.
 DR Growth factor; Cytokine; Glycoprotein; Signal.
 FT SIGNAL 1 19
 FT PROPEP 20 272
 FT CHAIN 273 395
 FT DISULFID 294 360
 FT DISULFID 323 392
 FT DISULFID 327 394
 FT DISULFID 359 359
 FT CARBOHYD 50 50
 FT CARBOHYD 116 116
 FT CARBOHYD 336 336
 FT SEQUENCE 395 AA; 43837 MW; 0FA3340DF5A6360E CRC64;
 Query Match 87.9%; Score 29; DB 1; Length 395;
 Best Local Similarity 80.0%; Pred. No. 78;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 323 CDGEC 327
 RESULT 32
 ID BMP7_XENLA STANDARD; PRT; 426 AA.
 AC P30866; O91645;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bone morphogenetic protein 7 precursor (BMP-7) (osteogenic protein 1)
 DE (OP-1) (XBMP7).
 GN BMP7.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;

OC Xenopodinae; Xenopus.
 OX NCBI_Taxid=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92378616; PubMed=1510675;
 RA Nishimatsu S., Suzuki A., Shoda A., Murakami K., Ueno N.;
 RT "Genes for bone morphogenetic proteins are differentially transcribed
 in early amphibian embryos.";
 RL Biochem. Biophys. Res. Commun. 186:1487-1495(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hawley S.H.B., Muntenberg-Stapleton K., Hashimoto C.,
 RA Laurent M.N., Watabe T., Blumberg B.W., Cho K.W.Y.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INDICES CARTILAGE AND BONE FORMATION. MAY BE THE
 OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF
 EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION
 AND BONE HOMEOSTASIS.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X63427; CAA45021.1; -.
 DR EMBL, U38559; AAA82616.1; -.
 DR PIR; JH0690; JH0690.
 DR HSSP; P18075; 1BMP.
 DR InterPro; IPR002400; GF_cyknoc.
 DR InterPro; IPR001839; TGFb.
 DR InterPro; IPR001111; TGFb_N.
 DR Pfam; PF000019; TGF-beta; 1.
 DR Pfam; PF00688; TGFb propeptide; 1.
 DR PRINTS; PRO0438; GFGYKNOT.
 DR ProDom; PD000357; TGFb; 1.
 DR SMART; SM00204; TGFb; 1.
 DR PROSITE; PS00250; TGF_BETA_1; 1.
 DR Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
 FT SIGNAL 1 22
 FT PROPEP 23 282
 FT CHAIN 283 426
 FT DISULFID 325 391
 FT DISULFID 354 423
 FT DISULFID 358 425
 FT DISULFID 390 390
 FT CARBOHYD 177 177
 FT CARBOHYD 307 307
 FT CARBOHYD 367 367
 FT CARBOHYD 317 319
 FT CONFLICT 355 355
 FT SEQUENCE 426 AA; 48965 MW; 6401D5151AC97117 CRC64;
 Query Match 87.9%; Score 29; DB 1; Length 426;
 Best Local Similarity 80.0%; Pred. No. 84;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 DB 354 CDGEC 358
 RESULT 33
 ID BMP5_MOUSE STANDARD; PRT; 452 AA.
 AC P49003;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)


```

DE Bone morphogenetic protein 5 precursor (BMP-5).
GN BMP5 OR BMP-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/KW;
RC MEDLINE=95046894; PubMed=7958439;
RA King J.A., Marker P.C., Seung K.J., Kingsley D.M.;
RT "BMP5 and the molecular, skeletal, and soft-tissue alterations in
RT short ear mice.";
RL Dev. Biol. 166:112-122(1994).
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L41145; AAA64612.1; -.
CC HSSP: P18075; 1BMP.
CC MGD: MGI:88181; BMP5.
CC InterPro: IPR002400; GF_cysknot.
CC InterPro: IPR001839; TGFb.
CC InterPro: IPR001111; TGFb_N.
CC Pfam: PF000019; TGF-beta; 1.
CC Pfam: PF00688; TGFb_propeptide; 1.
CC PRINTS: PR00438; GFGYSKNOT.
CC ProDom: PD000357; TGFb; 1.
CC SMART: SM00204; TGFb; 1.
CC PROSITE: PS00250; TGF_BETA_1; 1.
CC Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
CC SIGNAL 1 25 POTENTIAL.
CC PROPEP 26 320 POTENTIAL.
CC CHAIN 321 452 BONE MORPHOGENETIC PROTEIN 5.
CC DISULFID 351 417 BY SIMILARITY.
CC DISULFID 380 449 BY SIMILARITY.
CC DISULFID 384 451 BY SIMILARITY.
CC DISULFID 416 416 INTERCHAIN (BY SIMILARITY).
CC CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 452 AA; 51511 MW; AAD9521EC94A7BD5 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 452;
Best Local Similarity 80.0%; Pred. No. 88;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5
DB 380 CDGEC 384

RESULT 34
BMP5_HUMAN STANDARD; PRT; 454 AA.
ID P22003; Q9NTM5; Q9H547;
DT 01-AUG-1991 (Rel. 19, Created)
DE Bone morphogenetic protein 5 precursor (BMP-5).
GN BMP5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone;
RC MEDLINE=91086608; PubMed=2263636;
RA Celeste A.J., Yamazaki J.A., Taylor R.C., Hewick R.M., Rosen V.,
RA Wang E.A., Wozney J.M.;
RT "Identification of transforming growth factor beta family members
RT present in bone-inductive protein purified from bovine bone.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9843-9847(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Bates K., Tracey A.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE LUNGS AND LIVER.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M60314; AAA6736.1; -.
CC EMBL: A1337178; CAC12765.1; -.
CC EMBL: A133386; CAB1657.1; -.
CC PIR: A39263; A39263.
CC HSSP: P18075; 1BMP.
CC Genew: HGNC:1072; BMP5.
CC MIM: 112265; -.
CC InterPro: IPR002400; GF_cysknot.
CC InterPro: IPR001839; TGFb.
CC InterPro: IPR001111; TGFb_N.
CC Pfam: PF000019; TGF-beta; 1.
CC Pfam: PF00688; TGFb_propeptide; 1.
CC PRINTS: PR00438; GFGYSKNOT.
CC ProDom: PD000357; TGFb; 1.
CC SMART: SM00204; TGFb; 1.
CC PROSITE: PS00250; TGF_BETA_1; 1.
CC Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
CC SIGNAL 1 30 POTENTIAL.
CC PROPEP 31 322 POTENTIAL.
CC CHAIN 323 454 BONE MORPHOGENETIC PROTEIN 5.
CC DISULFID 353 419 BY SIMILARITY.
CC DISULFID 382 451 BY SIMILARITY.
CC DISULFID 386 453 BY SIMILARITY.
CC DISULFID 418 418 INTERCHAIN (BY SIMILARITY).
CC CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 454 AA; 51736 MW; 631277413CC22EE CRC64;

Query Match 87.9%; Score 29; DB 1; Length 454;
Best Local Similarity 80.0%; Pred. No. 88;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDABC 5
DB 382 CDGEC 386

RESULT 35
BMP6_MOUSE STANDARD; PRT; 510 AA.
ID P20722;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```


DE Bone morphogenetic protein 6 precursor (BMP-6) (VG-1-related protein)
 DE (VGR-1).
 DE BMP6 OR BMP-6 OR VGR1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=94375533; PubMed=8089189;
 RA Gitelman S.E., Kobrin M.S., Ye J.Q., Lopez A.R., Lee A., Derynck R.;
 RT "Recombinant Vgr-1/BMP-6-expressing tumors induce fibrosis and
 RL endochondral bone formation in vivo.";
 RL J. Cell Biol. 126:1595-1609(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97222480; PubMed=9069123;
 RA Gitelman S.E., Kobrin M., Lee A., Fet V., Lyons K., Hogan B.L.M.,
 RA Derynck R.;
 RL "Structure and sequence of the mouse Bmp6 gene.";
 RL Mamm. Genome 8:212-214(1997).
 [3]
 RP SEQUENCE OF 73-510 FROM N.A.
 RX MEDLINE=98282810; PubMed=2734307;
 RA Lyons K., Graycar J.L., Lee A., Hashmi S., Lindquist P.B., Chen E.Y.,
 RA Hogan B.L.M., Derynck R.;
 RT "Vgr-1, a mammalian gene related to Xenopus Vg-1, is a member of the
 RT transforming growth factor beta gene superfamily.";
 RT Proc. Natl. Acad. Sci. U.S.A. 86:4554-4558(1989).
 CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE LUNGS. LOW LEVELS SEEN IN
 CC THE KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X80992; CAAS6917.1; -;
 DR EMBL: U73520; AAB18235.1; -;
 DR EMBL: U73515; AAB18235.1; JOINED.
 DR EMBL: U73516; AAB18235.1; JOINED.
 DR EMBL: U73517; AAB18235.1; JOINED.
 DR EMBL: U73518; AAB18235.1; JOINED.
 DR EMBL: U73519; AAB18235.1; JOINED.
 DR EMBL: J04566; AAA40548.1; -;
 DR PIR: A33925; A33925.
 DR HSSP: P18075; 1BMP.
 DR WGD; MG1:88182; 1BMP.
 DR InterPro: IPR002400; GF_cysknot.
 DR InterPro: IPR001839; TGFb.
 DR InterPro: IPR001111; TGFb_N.
 DR Pfam: PF00688; TGF_beta; 1.
 DR Pfam: PF00688; TGF_beta; 1.
 DR PRINTS: PR00438; GFCYSKNOT.
 DR PRODOM: PD000357; TGFb; 1.
 DR SMART: SM00204; TGFb; 1.
 DR PROSITE: PS00250; TGF_BETA_1; 1.
 KW Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
 FT SIGNAL 1
 FT PROPEP 21 378
 FT CHAIN 379 510
 FT DISULFID 409 475
 FT DISULFID 438 507
 FT DISULFID 442 509
 FT DISULFID 474 474
 FT CARBOHYD 238 238
 FT 266 266
 N-LINKED (GLCNAC. .) (POTENTIAL).
 N-LINKED (GLCNAC. .) (POTENTIAL).
 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 383 383 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 451 451 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 73 73 K -> M (IN REF. 3).
 FT CONFLICT 75 75 E -> K (IN REF. 3).
 FT CONFLICT 86 86 L -> P (IN REF. 3).
 SQ SEQUENCE 510 AA; 56431 MW; 910B68531289FCD2 CRC64;
 Query Match 87.9%; Score 29; DB 1; Length 510;
 Best Local Similarity 80.0%; Pred. No. 97;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CDAC 5
 DB 438 CDACC 442
 RESULT 36
 BMD6_HUMAN STANDARD; PRT; 513 AA.
 ID BMD6_HUMAN
 AC P22004;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bone morphogenetic protein 6 precursor (BMP-6).
 GN BMP6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=bone;
 RX MEDLINE=91086608; PubMed=2263636;
 RA Celeste A.V., Iamazzi J.A., Taylor R.C., Hewick R.M., Rosen V.,
 RA Wang E.A., Wozney J.M.;
 RT "Identification of transforming growth factor beta family members
 RT present in bone-inductive protein purified from bovine bone.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9843-9847(1990).
 CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M60315; AAA6737.1; -;
 DR PIR: B39263; B39263.
 DR HSSP: P18075; 1BMP.
 DR Genew; HGNC:1073; BMP6.
 DR MIM; 112266; -;
 DR InterPro: IPR002400; GF_cysknot.
 DR InterPro: IPR001839; TGFb.
 DR InterPro: IPR001111; TGFb_N.
 DR Pfam: PF00688; TGF_beta; 1.
 DR Pfam: PF00688; TGF_beta; 1.
 DR PRINTS: PR00438; GFCYSKNOT.
 DR PRODOM: PD000357; TGFb; 1.
 DR SMART: SM00204; TGFb; 1.
 DR PROSITE: PS00250; TGF_BETA_1; 1.
 KW Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
 FT SIGNAL 1
 FT PROPEP 21 381
 FT CHAIN 382 513
 FT DISULFID 412 478
 FT DISULFID 441 510
 FT DISULFID 445 512
 FT DISULFID 477 477
 FT CARBOHYD 241 241
 N-LINKED (GLCNAC. .) (POTENTIAL).
 N-LINKED (GLCNAC. .) (POTENTIAL).
 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 513 AA; 57225 MW; 3F19155B86049278 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 513;
 Best Local Similarity 80.0%; Pred. No. 98;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDAC 5
 DB 441 CDGEC 445

RESULT 37
 TRA_BPMU STANDARD; PRT; 663 AA.
 ID TRA_BPMU STANDARD; PRT; 663 AA.
 C P07636; P06021; 07, Created)
 01-APR-1988 (Rel. 07, Created)
 01-FEB-1996 (Rel. 33, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transposase.
 GN A OR 3.
 OS Bacteriophage Mu.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC Mu-like viruses.
 OC NCBI_TaxID=10677;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86067968; Pubmed=2999776;
 RA Harsley R.M., Getzoff E.D., Baldwin D.L., Miller J.T., Chaconas G.;
 RT "Primary structure of phage mu transposase: homology to mu
 repressor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7676-7680(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Pries H., Brauer B., Schmidt C., Kamp D.;
 RT "Sequence of the left end of Mu.";
 RL (in) Symonds N., Toussein A., van de Putte P., Howe M.M. (eds.);
 RL phage Mu. pp.277-286, Cold Spring Harbor Laboratory Press,
 RL New York (1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Morgan G., Hatfull G., Hendrix R.;
 RT "Genome of bacteriophage Mu and comparison with the Haemophilus
 influenzae Mu-like prophage PluM.";
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-88 FROM N.A.
 RX MEDLINE=83012203; Pubmed=6214696;
 RA Pries H., Kamp D., Kahmann R., Brauer B., Delius H.;
 RT "Nucleotide sequence of the immunity region of bacteriophage Mu.";
 RL Mol. Gen. Genet. 186:315-321(1982).
 RN [5]
 RP SEQUENCE OF 1-84 FROM N.A.
 RX MEDLINE=83218562; Pubmed=6222246;
 RA Toussein A., Faalen M., Desmet L., Allet B.;
 RT "The products of gene A of the related phages Mu and D108 differ in
 their specificities.";
 RL Mol. Gen. Genet. 190:70-79(1983).
 RN [6]
 RP STRUCTURE BY NMR OF 1-76.
 RX MEDLINE=95187707; Pubmed=7881904;
 RA Glubb R.T., Omichinski J.G., Savilanti H., Mizuchi K.,
 RA Gronenborn A.M., Clore G.M.;
 RT "A novel class of winged helix-turn-helix protein: the DNA-binding
 domain of Mu transposase.";
 RL Structure 2:1041-1048(1994).
 RN [7]
 RP STRUCTURE BY NMR OF 76-174.
 RX MEDLINE=98070329; Pubmed=9405381;
 RA Schumacher S., Glubb R.T., Cai M., Mizuchi K., Clore G.M.,

RA Gronenborn A.M.;
 RT "Solution structure of the Mu end DNA-binding beta subdomain of
 phage Mu transposase: modular DNA recognition by two tethered
 RT domains.";
 RL EMBO J. 16:7532-7541(1997).
 RN [8]
 RP STRUCTURE BY NMR OF 173-247.
 RX MEDLINE=9805037; Pubmed=9367742;
 RA Glubb R.T., Schumacher S., Mizuchi K., Gronenborn A.M., Clore G.M.;
 RT "Solution structure of the I gamma subdomain of the Mu end
 RT DNA-binding domain of phage Mu transposase.";
 RL J. Mol. Biol. 273:19-25(1997).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 248-574.
 RX MEDLINE=95354202; Pubmed=7628012;
 RA Rice P., Mizuchi K.;
 RT "Structure of the bacteriophage Mu transposase core: a common
 RT structural motif for DNA transposition and retroviral integration.";
 RL Cell 82:209-220(1995).
 CC -1- FUNCTION: THIS TRANSPOSASE IS ESSENTIAL FOR INTEGRATION,
 CC REPLICATION-TRANSDUCTION, AND EXCISION OF MU DNA.
 CC -1- REPLICATION-TRANSDUCTION: MU CAN TRANSPOSE ITS DNA INTO MULTIPLE SITES IN
 CC MANY BACTERIAL GENOMES AND MEDIANE A VARIETY OF DNA
 CC REARRANGEMENTS. TRANSDUCTION REQUIRES BOTH TRANSPOSASE (ENCODED
 CC BY GENE A) AND TRANSDUCTION ENHANCER (ENCODED BY GENE B).
 CC -1- MISCELLANEOUS: UNLIKE OTHER TRANSPOSASONS MU HAS DISSIMILAR
 CC SEQUENCES AT ITS LEFT AND RIGHT ENDS. TRANSPOSASE APPARENTLY BINDS
 CC 3 SPECIFIC BLOCKS OF SEQUENCES AT EACH END OF MU DNA.
 CC -1- MISCELLANEOUS: THE A GENE IS REGULATED BY THE REPRESSOR C, WHICH
 CC BINDS TO AN OPERATOR SEQUENCE & TURNS OFF TRANSCRIPTION. REPRESSOR
 CC ON MU ENDS AS THE TRANSPOSASE, AND TRANSPOSASE CAN BIND TO
 CC FRAGMENTS CONTAINING THE MU OPERATOR SEQUENCE.
 CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE H1478.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

EMBL; M1195; AAA32369.1; -
 EMBL; M64097; AAA32379.1; -
 EMBL; AF083977; AAF01083.1; -
 EMBL; V01464; CAA24713.1; -
 EMBL; V00868; CAA24236.1; -
 PIR; A24746; TQBPV.
 PDB; 1TNS; 14-FEB-95.
 PDB; 1TNT; 14-FEB-95.
 PDB; 1BCN; 15-OCT-95.
 PDB; 1BCO; 15-OCT-95.
 PDB; 2E2H; 03-DEC-97.
 PDB; 2E2I; 03-DEC-97.
 PDB; 2E2J; 14-JAN-98.
 PDB; 2E2L; 14-JAN-98.
 InterPro: IPR003314; Mu DNA bind.
 InterPro: IPR004189; Mu transposase.
 Pfam: PF02316; Mu_DNA_bind; 1.
 Pfam: PF02914; Mu_transposase; 1.
 KX DNA integration; Transposase element; DNA excision;
 KM DNA integration; Transposase element; DNA excision;
 FT DNA_BIND 35 55
 FT DNA_BIND 360 409 H-T-H MOTIF (POTENTIAL).
 FT CONFLICT 66 66 G -> R (IN REF. 5).
 FT CONFLICT 408 408 P -> S (IN REF. 2).
 SO SEQUENCE 663 AA; 75003 MW; B882CFCBFC0B2E3 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 663;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CDAC 5
Db 555 CEAC 559

RESULT 38
YAGI_SCHPO

ID YAGI_SCHPO STANDARD; PRT; 905 AA.

AC 009866; Q9UHU9;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein C12G12.01c in chromosome I.
SPAC12G12.01c OR SPAC630.02.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetes; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
[1]

SEQUENCE FROM N.A.

STRAIN=972;

MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajadream W.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.D., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moul S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakert G., Aert R., Robben J., Grymptre B.,
RA Woldens I., Vastakert E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Beer P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Carruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Uesely D., Barrell B.G., Nurse P.,
RA The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880(2002).
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: TO YEAST YBL104C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AL109832; CAB52724.1; -
DR EMBL; Z66568; CA91496.1; -
DR InterPro; IPR001230; Ptenyl_site.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 2.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00678; WD REPEATS_1; FALSE NEG.
DR PROSITE; PS00682; WD REPEATS_2; FALSE NEG.
DR PROSITE; PS00294; WD REPEATS_REGION; FALSE NEG.
KW Hypothetical protein; Repeat; WD repeat.
FT REPEAT 42 92 WD 1.
FT REPEAT 86 128 WD 2.
FT REPEAT 136 175 WD 3.

FT REPEAT 177 217 WD 4.
SQ SEQUENCE 905 AA; 101948 MW; 56BCF35CA07F709B CRC64;

Query Match 87.9%; Score 29; DB 1; Length 905;

Best Local Similarity 80.0%; Pred. No. 1.6e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CDAC 5
Db 898 CDCRC 902

RESULT 39

ID BAR3_CHITE STANDARD; PRT; 1700 AA.

AC 003376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Balbiant ring protein 3 precursor.
GN BR3.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
NCBI_TaxID=7153;
[1]

SEQUENCE FROM N.A.

TISSUE=Salivary gland;

MEDLINE=9012404; PubMed=1689777;

RA Paulsen G., Lendahl U., Galli J., Ericsson C., Wieslander L.,

RT "The Balbiant ring 3 gene in Chironomus tentans has a diverged

RT repetitive structure split by many introns."

RT J. Mol. Biol. 211:331-349(1990).

CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR

CC STRUCTURE, THE LARVAL TUBE. BALBIANT RING PROTEIN 3 COULD PLAY A

CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS

CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE

CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.

CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X52263; CA936506.1; -
DR PIR; S08167; S08167.
DR HSP; P15358; ISKZ.

DR InterPro; IPR004153; CXKC_repeat.

DR Pfam; PF03128; CXKC; 71.

DR Repeat; Signal.

FT SIGNAL 1 20

FT CHAIN 21 1700

SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 1700;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CDAC 5
Db 680 CDCRC 684

RESULT 40
SIX2_BUTAR STANDARD; PRT; 61 AA.

AC P80962;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Depressant insect toxin 2 (Bait2).
 OS Butacus arenicola (North African scorpion).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
 OC Butchoidea; Butidae; Butacus.
 OX NCBI_TaxID=70335;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RX MEDLINE=97182587; PubMed=9030726;
 RA Cestele S., Kopeyan C., Oughident R., Mansuelle P., Granier C.,
 RA Rochat H.;
 RT "Biochemical and pharmacological characterization of a depressant
 insect toxin from the venom of the scorpion Butacus arenicola.";
 RL Eur. J. Biochem. 243:93-99(1997).
 -1- FUNCTION: BINDS TO SODIUM CHANNELS AND INHIBITS THE INACTIVATION
 OF THE ACTIVATED CHANNELS, THEREBY BLOCKING NEURONAL TRANSMISSION.
 THIS TOXIN INDUCES A SLOW, DEPRESSANT, FLACCID PARALYSIS ON FLY
 LARVAE. IT IS ACTIVE ONLY ON INSECTS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.
 CC ALPHA-TOXIN SUBFAMILY.
 CC HSSP; P01494; 28N3.
 DR InterPro; IPR003614; Kncot1.
 DR InterPro; IPR001219; Neurotoxin.
 DR InterPro; IPR002061; Scorpion_toxinL.
 DR Pfam; PF00537; toxin_3; 1.
 DR PRINTS; PR00284; TOXIN.
 DR ProDom; PD000908; Scorpion_toxinL; 1.
 DR SMART; SM00505; Kncot1; 1.
 KM Neurotoxin; Sodium channel inhibitor.
 FT DISULFID 10 60 BY SIMILARITY.
 FT DISULFID 14 35 BY SIMILARITY.
 FT DISULFID 21 42 BY SIMILARITY.
 FT DISULFID 25 44 BY SIMILARITY.
 SO SEQUENCE 61 AA; 6845 MW; 4143435560DCAFB CRC64;

Query Match 84.8%; Score 28; DB 1; Length 61;
 Best Local Similarity 80.0%; Pred. No. 25;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDAEC 5.
 21 CDKEC 25


```

GN EI.
OS Highlands J virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11024;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC-431;
RA Clinis M.J., Kang, W, Weaver S.C.;
RT "Genetic Conservation of Highlands J Viruses.";
RL Virology 218:0-0(1996).
DR EMBL; U52576; AAB02206.1; -.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
FT NON TER
SQ SEQUENCE 348 AA; 37353 MW; 10E406B56406EE32 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 285 CDAEC 289

RESULT 3
ID 080739 PRELIMINARY; PRT; 348 AA.
AC 080739;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN EI.
OS Highlands J virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11024;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-8-74;
RA Clinis M.J., Kang, W, Weaver S.C.;
RT "Genetic Conservation of Highlands J Viruses.";
RL Virology 218:0-0(1996).
DR EMBL; U52580; AAB02210.1; -.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
FT NON TER
SQ SEQUENCE 348 AA; 37395 MW; 18468C9DD52102CC CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 285 CDAEC 289

RESULT 4
ID 080740 PRELIMINARY; PRT; 348 AA.
AC 080740;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN EI.
OS Highlands J virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11024;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=73V2540;
RA Clinis M.J., Kang, W, Weaver S.C.;
RT "Genetic Conservation of Highlands J Viruses.";
RL Virology 218:0-0(1996).
DR EMBL; U52581; AAB02211.1; -.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
FT NON TER
SQ SEQUENCE 348 AA; 37339 MW; EED9EB5D95D8550F CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 285 CDAEC 289

RESULT 5
ID 080741 PRELIMINARY; PRT; 348 AA.
AC 080741;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN EI.
OS Highlands J virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11024;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCU5-6X;
RA Clinis M.J., Kang, W, Weaver S.C.;
RT "Genetic Conservation of Highlands J Viruses.";
RL Virology 218:0-0(1996).
DR EMBL; U52582; AAB02212.1; -.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
FT NON TER
SQ SEQUENCE 348 AA; 37397 MW; 86DAEF7AD88AD77F CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
DB 285 CDAEC 289

RESULT 6
ID 080742 PRELIMINARY; PRT; 348 AA.
AC 080742;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN EI.
OS Highlands J virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11024;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WX3-2AP;
RA Clinis M.J., Kang, W, Weaver S.C.;
RT "Genetic Conservation of Highlands J Viruses.";

```

RL Virology 218:0-0(1996).
 DR EMBL; U52584; AAB02214.1; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 FT NON_TER 1
 SQ SEQUENCE 348 AA; 37381 MW; 045A1D54DA01BC99 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
 DB 285 CDAC 289

RESULT 7
 ID 080743 PRELIMINARY; PRT; 348 AA.
 AC 080743;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN E1.
 OS Highlands J virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 RX NCB1_TaxID=11024;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SP72-666;
 RA Clinis M.J., Kang, W, Weaver S.C.;
 RT "Genetic Conservation of Highlands J Viruses."
 RL Virology 218:0-0(1996).
 DR EMBL; U52585; AAB02215.1; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 FT NON_TER 1
 SQ SEQUENCE 348 AA; 37436 MW; 10E5DDC5FA6B5932 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
 DB 285 CDAC 289

RESULT 8
 ID 080744 PRELIMINARY; PRT; 348 AA.
 AC 080744;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN E1.
 OS Highlands J virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 RX NCB1_TaxID=11024;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=3-81690;
 RA Clinis M.J., Kang, W, Weaver S.C.;
 RT "Genetic Conservation of Highlands J Viruses."
 RL Virology 218:0-0(1996).
 DR EMBL; U52588; AAB02218.1; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 FT NON_TER 1

SQ SEQUENCE 348 AA; 37297 MW; FEE49C58F4071F92 CRC64;
 Query Match 100.0%; Score 33; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
 DB 285 CDAC 289

RESULT 9
 ID 080745 PRELIMINARY; PRT; 348 AA.
 AC 080745;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN E1.
 OS Highlands J virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 RX NCB1_TaxID=11024;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=R362;
 RA Clinis M.J., Kang, W, Weaver S.C.;
 RT "Genetic Conservation of Highlands J Viruses."
 RL Virology 218:0-0(1996).
 DR EMBL; U52589; AAB02219.1; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 FT NON_TER 1
 SQ SEQUENCE 348 AA; 37323 MW; D256E225A2ABD866 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
 DB 285 CDAC 289

RESULT 10
 ID 080746 PRELIMINARY; PRT; 348 AA.
 AC 080746;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN E1.
 OS Highlands J virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 RX NCB1_TaxID=11024;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=F25;
 RA Clinis M.J., Kang, W, Weaver S.C.;
 RT "Genetic Conservation of Highlands J Viruses."
 RL Virology 218:0-0(1996).
 DR EMBL; U52590; AAB02220.1; -
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 FT NON_TER 1
 SQ SEQUENCE 348 AA; 37353 MW; 10E516A4AF7AC742 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 285 CDAC 289

RESULT 11

Q90205 PRELIMINARY; PRT; 348 AA.
AC 090205
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN EI.
OS Highlands J virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
NCBI_TaxID=11024;

SEQUENCE FROM N.A.
MEDLINE=96193746; PubMed=8610461;
[1]
RA Clinis M.J., Kang W., Weaver S.C.;
RT "Genetic conservation of Highlands J viruses.";
RL Virology 218:343-351 (1996).
DR EMBL; U52575; AAB02205.1; -
DR EMBL; U52577; AAB02207.1; -
DR EMBL; U52578; AAB02208.1; -
DR EMBL; U52579; AAB02209.1; -
DR EMBL; U52583; AAB02213.1; -
DR EMBL; U52586; AAB02216.1; -
DR EMBL; U52591; AAB02221.1; -
DR EMBL; U52592; AAB02222.1; -
DR EMBL; U52593; AAB02223.1; -
DR InterPro; IPR002546; Alpha_E1_glycop.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
KM Envelope protein.
FT NON TER 1
SQ SEQUENCE 348 AA; 37367 MW; 10E406B5641A1C32 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 348;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 285 CDAC 289

RESULT 12
Q87ZC9 PRELIMINARY; PRT; 352 AA.
AC 087ZC9
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE RNA phosphate cyclase.
GN RCL1 OR MK0001.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
NCBI_TaxID=2320;

SEQUENCE FROM N.A.
STRAIN=AV19 / DSM 6324 / JCM 9639;
MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Natcharbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,
RA Malynh A.G., Koonin E.V., Kozlyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
DR EMBL; AE010302; AAM01218.1; -
KM Complete proteome.
SQ SEQUENCE 352 AA; 38092 MW; 9E0952828B031E5A CRC64;

Query Match 100.0%; Score 33; DB 17; Length 352;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 61 CDAC 65

RESULT 13

Q9W188 PRELIMINARY; PRT; 735 AA.
AC 09W188
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG3376 protein (LD42214P).
GN CG3376.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;

SEQUENCE FROM N.A.
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Asgavani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kappen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirakae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chavez M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A6003463; AAF47189.1; -
 DR EMBL: AY061466; AAL29014.1; -
 DR FlyBase: FBgn003497; CG3376.
 DR InterPro: IPR004843; M-peptidase.
 DR InterPro: IPR004844; S/T phosphatase.
 DR InterPro: IPR000004; SAPB.
 DR Pfam: PF00149; Metallophos; 1.
 DR SMART: SM00118; SAPB; 1.
 SQ SEQUENCE 735 AA; 84183 MW; A507843469378F5 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 735;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CDAEC 5
 |||||
 658 CDAEC 662

RESULT 14
 ID 09VU17 PRELIMINARY; PRT; 744 AA.
 AC 09VU17;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DE CG3349 protein.
 GN CG3349
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abrial J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunker B.C., Dunn P.,
 RA Durbin K.J., Ewing L.C., Ferraz C., Ferrara S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalili M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklav G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palczolo M., Pitchen G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Maassman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: A6003532; AAF49690.1; -
 DR FlyBase: FBgn003459; CG3349.
 SQ SEQUENCE 744 AA; 86456 MW; 896CF66879D3A45 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 744;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CDAEC 5
 |||||
 716 CDAEC 720

RESULT 15
 ID 08S2S2 PRELIMINARY; PRT; 1376 AA.
 AC 08S2S2;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
 DE IDJ0182P.
 GN FUR2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chavez M., Chavez C., Dorsett V., Drensek D., Farfan D., Frise B.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY070553; AAL48024.1;
 SQ SEQUENCE 1376 AA; 149716 MW; B6704BA89A3A88FB CRC64;

Query Match 100.0%; Score 33; DB 5; Length 1376;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CDAEC 5
 |||||
 664 CDAEC 668

RESULT 16
 ID 024301 PRELIMINARY; PRT; 1679 AA.
 AC 024301;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DE FUR2 protein.
 GN FUR2 OR DFUR2 OR CG4235 OR CG18734.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner-M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bandaru D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jellali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclob J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ISO-1;
 RC MEDLINE=92381036; PubMed=1512259;
 RA Roebroek A.J., Creemers J.W., Pauli I.G., Kurzik-Dumke U., Rentrop M.,
 RA Gattef E.A., Leunissen J.A., de Ven W.J.;
 RT "Cloning and functional expression of Dfurn2, a subtilisin-like
 protease processing enzyme of *Drosophila melanogaster* with multiple
 repeats of a cysteine motif.";
 RL J. Biol. Chem. 267:17208-17215 (1992).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ISO-1;
 RC MEDLINE=95186060; PubMed=7880443;
 RA Roebroek A.J., Ayoubi T.A., Creemers J.W., Pauli I.G., Ven W.J.;
 RT "The Dfurn2 gene of *Drosophila melanogaster*: genetic organization,
 expression during embryogenesis, and pro-protein processing activity
 of its translational product Dfurn2.";
 RL DNA Cell Biol. 14:223-234 (1995).
 DR EMBL: A6003502; A6048598.1; -;
 DR EMBL: L33831; A6048598.1; -;
 DR HSSP: Q99405; 1MPT.
 DR MEROPS: S08.049; -;
 DR FLYBASE: FBgn004598; Fur2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF01483; P_1.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PRODOM: PD000717; P_domain; 1.
 DR SMART: SM00181; EGF; 1.

DR SMART: SM00261; FU; 10.
 DR PROSITE: PS00136; SUBTILASE ASP; 1.
 DR PROSITE: PS00137; SUBTILASE HIS; 1.
 DR PROSITE: PS00138; SUBTILASE SER; 1.
 DR PROSITE: PS00139; SUBTILASE TRP; 1.
 DR PROSITE: PS00140; SUBTILASE VAL; 1.
 DR PROSITE: PS00141; SUBTILASE THR; 1.
 DR PROSITE: PS00142; SUBTILASE LEU; 1.
 DR PROSITE: PS00143; SUBTILASE ILE; 1.
 DR PROSITE: PS00144; SUBTILASE MET; 1.
 DR PROSITE: PS00145; SUBTILASE PHE; 1.
 DR PROSITE: PS00146; SUBTILASE TYR; 1.
 DR PROSITE: PS00147; SUBTILASE ASP; 1.
 DR PROSITE: PS00148; SUBTILASE HIS; 1.
 DR PROSITE: PS00149; SUBTILASE SER; 1.
 DR PROSITE: PS00150; SUBTILASE TRP; 1.
 DR PROSITE: PS00151; SUBTILASE VAL; 1.
 DR PROSITE: PS00152; SUBTILASE THR; 1.
 DR PROSITE: PS00153; SUBTILASE LEU; 1.
 DR PROSITE: PS00154; SUBTILASE ILE; 1.
 DR PROSITE: PS00155; SUBTILASE MET; 1.
 DR PROSITE: PS00156; SUBTILASE PHE; 1.
 DR PROSITE: PS00157; SUBTILASE TYR; 1.
 DR PROSITE: PS00158; SUBTILASE ASP; 1.
 DR PROSITE: PS00159; SUBTILASE HIS; 1.
 DR PROSITE: PS00160; SUBTILASE SER; 1.
 DR PROSITE: PS00161; SUBTILASE TRP; 1.
 DR PROSITE: PS00162; SUBTILASE VAL; 1.
 DR PROSITE: PS00163; SUBTILASE THR; 1.
 DR PROSITE: PS00164; SUBTILASE LEU; 1.
 DR PROSITE: PS00165; SUBTILASE ILE; 1.
 DR PROSITE: PS00166; SUBTILASE MET; 1.
 DR PROSITE: PS00167; SUBTILASE PHE; 1.
 DR PROSITE: PS00168; SUBTILASE TYR; 1.
 DR PROSITE: PS00169; SUBTILASE ASP; 1.
 DR PROSITE: PS00170; SUBTILASE HIS; 1.
 DR PROSITE: PS00171; SUBTILASE SER; 1.
 DR PROSITE: PS00172; SUBTILASE TRP; 1.
 DR PROSITE: PS00173; SUBTILASE VAL; 1.
 DR PROSITE: PS00174; SUBTILASE THR; 1.
 DR PROSITE: PS00175; SUBTILASE LEU; 1.
 DR PROSITE: PS00176; SUBTILASE ILE; 1.
 DR PROSITE: PS00177; SUBTILASE MET; 1.
 DR PROSITE: PS00178; SUBTILASE PHE; 1.
 DR PROSITE: PS00179; SUBTILASE TYR; 1.
 DR PROSITE: PS00180; SUBTILASE ASP; 1.
 DR PROSITE: PS00181; SUBTILASE HIS; 1.
 DR PROSITE: PS00182; SUBTILASE SER; 1.
 DR PROSITE: PS00183; SUBTILASE TRP; 1.
 DR PROSITE: PS00184; SUBTILASE VAL; 1.
 DR PROSITE: PS00185; SUBTILASE THR; 1.
 DR PROSITE: PS00186; SUBTILASE LEU; 1.
 DR PROSITE: PS00187; SUBTILASE ILE; 1.
 DR PROSITE: PS00188; SUBTILASE MET; 1.
 DR PROSITE: PS00189; SUBTILASE PHE; 1.
 DR PROSITE: PS00190; SUBTILASE TYR; 1.
 DR PROSITE: PS00191; SUBTILASE ASP; 1.
 DR PROSITE: PS00192; SUBTILASE HIS; 1.
 DR PROSITE: PS00193; SUBTILASE SER; 1.
 DR PROSITE: PS00194; SUBTILASE TRP; 1.
 DR PROSITE: PS00195; SUBTILASE VAL; 1.
 DR PROSITE: PS00196; SUBTILASE THR; 1.
 DR PROSITE: PS00197; SUBTILASE LEU; 1.
 DR PROSITE: PS00198; SUBTILASE ILE; 1.
 DR PROSITE: PS00199; SUBTILASE MET; 1.
 DR PROSITE: PS00200; SUBTILASE PHE; 1.
 DR PROSITE: PS00201; SUBTILASE TYR; 1.
 DR PROSITE: PS00202; SUBTILASE ASP; 1.
 DR PROSITE: PS00203; SUBTILASE HIS; 1.
 DR PROSITE: PS00204; SUBTILASE SER; 1.
 DR PROSITE: PS00205; SUBTILASE TRP; 1.
 DR PROSITE: PS00206; SUBTILASE VAL; 1.
 DR PROSITE: PS00207; SUBTILASE THR; 1.
 DR PROSITE: PS00208; SUBTILASE LEU; 1.
 DR PROSITE: PS00209; SUBTILASE ILE; 1.
 DR PROSITE: PS00210; SUBTILASE MET; 1.
 DR PROSITE: PS00211; SUBTILASE PHE; 1.
 DR PROSITE: PS00212; SUBTILASE TYR; 1.
 DR PROSITE: PS00213; SUBTILASE ASP; 1.
 DR PROSITE: PS00214; SUBTILASE HIS; 1.
 DR PROSITE: PS00215; SUBTILASE SER; 1.
 DR PROSITE: PS00216; SUBTILASE TRP; 1.
 DR PROSITE: PS00217; SUBTILASE VAL; 1.
 DR PROSITE: PS00218; SUBTILASE THR; 1.
 DR PROSITE: PS00219; SUBTILASE LEU; 1.
 DR PROSITE: PS00220; SUBTILASE ILE; 1.
 DR PROSITE: PS00221; SUBTILASE MET; 1.
 DR PROSITE: PS00222; SUBTILASE PHE; 1.
 DR PROSITE: PS00223; SUBTILASE TYR; 1.
 DR PROSITE: PS00224; SUBTILASE ASP; 1.
 DR PROSITE: PS00225; SUBTILASE HIS; 1.
 DR PROSITE: PS00226; SUBTILASE SER; 1.
 DR PROSITE: PS00227; SUBTILASE TRP; 1.
 DR PROSITE: PS00228; SUBTILASE VAL; 1.
 DR PROSITE: PS00229; SUBTILASE THR; 1.
 DR PROSITE: PS00230; SUBTILASE LEU; 1.
 DR PROSITE: PS00231; SUBTILASE ILE; 1.
 DR PROSITE: PS00232; SUBTILASE MET; 1.
 DR PROSITE: PS00233; SUBTILASE PHE; 1.
 DR PROSITE: PS00234; SUBTILASE TYR; 1.
 DR PROSITE: PS00235; SUBTILASE ASP; 1.
 DR PROSITE: PS00236; SUBTILASE HIS; 1.
 DR PROSITE: PS00237; SUBTILASE SER; 1.
 DR PROSITE: PS00238; SUBTILASE TRP; 1.
 DR PROSITE: PS00239; SUBTILASE VAL; 1.
 DR PROSITE: PS00240; SUBTILASE THR; 1.
 DR PROSITE: PS00241; SUBTILASE LEU; 1.
 DR PROSITE: PS00242; SUBTILASE ILE; 1.
 DR PROSITE: PS00243; SUBTILASE MET; 1.
 DR PROSITE: PS00244; SUBTILASE PHE; 1.
 DR PROSITE: PS00245; SUBTILASE TYR; 1.
 DR PROSITE: PS00246; SUBTILASE ASP; 1.
 DR PROSITE: PS00247; SUBTILASE HIS; 1.
 DR PROSITE: PS00248; SUBTILASE SER; 1.
 DR PROSITE: PS00249; SUBTILASE TRP; 1.
 DR PROSITE: PS00250; SUBTILASE VAL; 1.
 DR PROSITE: PS00251; SUBTILASE THR; 1.
 DR PROSITE: PS00252; SUBTILASE LEU; 1.
 DR PROSITE: PS00253; SUBTILASE ILE; 1.
 DR PROSITE: PS00254; SUBTILASE MET; 1.
 DR PROSITE: PS00255; SUBTILASE PHE; 1.
 DR PROSITE: PS00256; SUBTILASE TYR; 1.
 DR PROSITE: PS00257; SUBTILASE ASP; 1.
 DR PROSITE: PS00258; SUBTILASE HIS; 1.
 DR PROSITE: PS00259; SUBTILASE SER; 1.
 DR PROSITE: PS00260; SUBTILASE TRP; 1.
 DR PROSITE: PS00261; SUBTILASE VAL; 1.
 DR PROSITE: PS00262; SUBTILASE THR; 1.
 DR PROSITE: PS00263; SUBTILASE LEU; 1.
 DR PROSITE: PS00264; SUBTILASE ILE; 1.
 DR PROSITE: PS00265; SUBTILASE MET; 1.
 DR PROSITE: PS00266; SUBTILASE PHE; 1.
 DR PROSITE: PS00267; SUBTILASE TYR; 1.
 DR PROSITE: PS00268; SUBTILASE ASP; 1.
 DR PROSITE: PS00269; SUBTILASE HIS; 1.
 DR PROSITE: PS00270; SUBTILASE SER; 1.
 DR PROSITE: PS00271; SUBTILASE TRP; 1.
 DR PROSITE: PS00272; SUBTILASE VAL; 1.
 DR PROSITE: PS00273; SUBTILASE THR; 1.
 DR PROSITE: PS00274; SUBTILASE LEU; 1.
 DR PROSITE: PS00275; SUBTILASE ILE; 1.
 DR PROSITE: PS00276; SUBTILASE MET; 1.
 DR PROSITE: PS00277; SUBTILASE PHE; 1.
 DR PROSITE: PS00278; SUBTILASE TYR; 1.
 DR PROSITE: PS00279; SUBTILASE ASP; 1.
 DR PROSITE: PS00280; SUBTILASE HIS; 1.
 DR PROSITE: PS00281; SUBTILASE SER; 1.
 DR PROSITE: PS00282; SUBTILASE TRP; 1.
 DR PROSITE: PS00283; SUBTILASE VAL; 1.
 DR PROSITE: PS00284; SUBTILASE THR; 1.
 DR PROSITE: PS00285; SUBTILASE LEU; 1.
 DR PROSITE: PS00286; SUBTILASE ILE; 1.
 DR PROSITE: PS00287; SUBTILASE MET; 1.
 DR PROSITE: PS00288; SUBTILASE PHE; 1.
 DR PROSITE: PS00289; SUBTILASE TYR; 1.
 DR PROSITE: PS00290; SUBTILASE ASP; 1.
 DR PROSITE: PS00291; SUBTILASE HIS; 1.
 DR PROSITE: PS00292; SUBTILASE SER; 1.
 DR PROSITE: PS00293; SUBTILASE TRP; 1.
 DR PROSITE: PS00294; SUBTILASE VAL; 1.
 DR PROSITE: PS00295; SUBTILASE THR; 1.
 DR PROSITE: PS00296; SUBTILASE LEU; 1.
 DR PROSITE: PS00297; SUBTILASE ILE; 1.
 DR PROSITE: PS00298; SUBTILASE MET; 1.
 DR PROSITE: PS00299; SUBTILASE PHE; 1.
 DR PROSITE: PS00300; SUBTILASE TYR; 1.
 DR PROSITE: PS00301; SUBTILASE ASP; 1.
 DR PROSITE: PS00302; SUBTILASE HIS; 1.
 DR PROSITE: PS00303; SUBTILASE SER; 1.
 DR PROSITE: PS00304; SUBTILASE TRP; 1.
 DR PROSITE: PS00305; SUBTILASE VAL; 1.
 DR PROSITE: PS00306; SUBTILASE THR; 1.
 DR PROSITE: PS00307; SUBTILASE LEU; 1.
 DR PROSITE: PS00308; SUBTILASE ILE; 1.
 DR PROSITE: PS00309; SUBTILASE MET; 1.
 DR PROSITE: PS00310; SUBTILASE PHE; 1.
 DR PROSITE: PS00311; SUBTILASE TYR; 1.
 DR PROSITE: PS00312; SUBTILASE ASP; 1.
 DR PROSITE: PS00313; SUBTILASE HIS; 1.
 DR PROSITE: PS00314; SUBTILASE SER; 1.
 DR PROSITE: PS00315; SUBTILASE TRP; 1.
 DR PROSITE: PS00316; SUBTILASE VAL; 1.
 DR PROSITE: PS00317; SUBTILASE THR; 1.
 DR PROSITE: PS00318; SUBTILASE LEU; 1.
 DR PROSITE: PS00319; SUBTILASE ILE; 1.
 DR PROSITE: PS00320; SUBTILASE MET; 1.
 DR PROSITE: PS00321; SUBTILASE PHE; 1.
 DR PROSITE: PS00322; SUBTILASE TYR; 1.
 DR PROSITE: PS00323; SUBTILASE ASP; 1.
 DR PROSITE: PS00324; SUBTILASE HIS; 1.
 DR PROSITE: PS00325; SUBTILASE SER; 1.
 DR PROSITE: PS00326; SUBTILASE TRP; 1.
 DR PROSITE: PS00327; SUBTILASE VAL; 1.
 DR PROSITE: PS00328; SUBTILASE THR; 1.
 DR PROSITE: PS00329; SUBTILASE LEU; 1.
 DR PROSITE: PS00330; SUBTILASE ILE; 1.
 DR PROSITE: PS00331; SUBTILASE MET; 1.
 DR PROSITE: PS00332; SUBTILASE PHE; 1.
 DR PROSITE: PS00333; SUBTILASE TYR; 1.
 DR PROSITE: PS00334; SUBTILASE ASP; 1.
 DR PROSITE: PS00335; SUBTILASE HIS; 1.
 DR PROSITE: PS00336; SUBTILASE SER; 1.
 DR PROSITE: PS00337; SUBTILASE TRP; 1.
 DR PROSITE: PS00338; SUBTILASE VAL; 1.
 DR PROSITE: PS00339; SUBTILASE THR; 1.
 DR PROSITE: PS00340; SUBTILASE LEU; 1.
 DR PROSITE: PS00341; SUBTILASE ILE; 1.
 DR PROSITE: PS00342; SUBTILASE MET; 1.
 DR PROSITE: PS00343; SUBTILASE PHE; 1.
 DR PROSITE: PS00344; SUBTILASE TYR; 1.
 DR PROSITE: PS00345; SUBTILASE ASP; 1.
 DR PROSITE: PS00346; SUBTILASE HIS; 1.
 DR PROSITE: PS00347; SUBTILASE SER; 1.
 DR PROSITE: PS00348; SUBTILASE TRP; 1.
 DR PROSITE: PS00349; SUBTILASE VAL; 1.
 DR PROSITE: PS00350; SUBTILASE THR; 1.
 DR PROSITE: PS00351; SUBTILASE LEU; 1.
 DR PROSITE: PS00352; SUBTILASE ILE; 1.
 DR PROSITE: PS00353; SUBTILASE MET; 1.
 DR PROSITE: PS00354; SUBTILASE PHE; 1.
 DR PROSITE: PS00355; SUBTILASE TYR; 1.
 DR PROSITE: PS00356; SUBTILASE ASP; 1.
 DR PROSITE: PS00357; SUBTILASE HIS; 1.
 DR PROSITE: PS00358; SUBTILASE SER; 1.
 DR PROSITE: PS00359; SUBTILASE TRP; 1.
 DR PROSITE: PS00360; SUBTILASE VAL; 1.
 DR PROSITE: PS00361; SUBTILASE THR; 1.
 DR PROSITE: PS00362; SUBTILASE LEU; 1.
 DR PROSITE: PS00363; SUBTILASE ILE; 1.
 DR PROSITE: PS00364; SUBTILASE MET; 1.
 DR PROSITE: PS00365; SUBTILASE PHE; 1.
 DR PROSITE: PS00366; SUBTILASE TYR; 1.
 DR PROSITE: PS00367; SUBTILASE ASP; 1.
 DR PROSITE: PS00368; SUBTILASE HIS; 1.
 DR PROSITE: PS00369; SUBTILASE SER; 1.
 DR PROSITE: PS00370; SUBTILASE TRP; 1.
 DR PROSITE: PS00371; SUBTILASE VAL; 1.
 DR PROSITE: PS00372; SUBTILASE THR; 1.
 DR PROSITE: PS00373; SUBTILASE LEU; 1.
 DR PROSITE: PS00374; SUBTILASE ILE; 1.
 DR PROSITE: PS00375; SUBTILASE MET; 1.
 DR PROSITE: PS00376; SUBTILASE PHE; 1.
 DR PROSITE: PS00377; SUBTILASE TYR; 1.
 DR PROSITE: PS00378; SUBTILASE ASP; 1.
 DR PROSITE: PS00379; SUBTILASE HIS; 1.
 DR PROSITE: PS00380; SUBTILASE SER; 1.
 DR PROSITE: PS00381; SUBTILASE TRP; 1.
 DR PROSITE: PS00382; SUBTILASE VAL; 1.
 DR PROSITE: PS00383; SUBTILASE THR; 1.
 DR PROSITE: PS00384; SUBTILASE LEU; 1.
 DR PROSITE: PS00385; SUBTILASE ILE; 1.
 DR PROSITE: PS00386; SUBTILASE MET; 1.
 DR PROSITE: PS00387; SUBTILASE PHE; 1.
 DR PROSITE: PS00388; SUBTILASE TYR; 1.
 DR PROSITE: PS00389; SUBTILASE ASP; 1.
 DR PROSITE: PS00390; SUBTILASE HIS; 1.
 DR PROSITE: PS00391; SUBTILASE SER; 1.
 DR PROSITE: PS00392; SUBTILASE TRP; 1.
 DR PROSITE: PS00393; SUBTILASE VAL; 1.
 DR PROSITE: PS00394; SUBTILASE THR; 1.
 DR PROSITE: PS00395; SUBTILASE LEU; 1.
 DR PROSITE: PS00396; SUBTILASE ILE; 1.
 DR PROSITE: PS00397; SUBTILASE MET; 1.
 DR PROSITE: PS00398; SUBTILASE PHE; 1.
 DR PROSITE: PS00399; SUBTILASE TYR; 1.
 DR PROSITE: PS00400; SUBTILASE ASP; 1.
 DR PROSITE: PS00401; SUBTILASE HIS; 1.
 DR PROSITE: PS00402; SUBTILASE SER; 1.
 DR PROSITE: PS00403; SUBTILASE TRP; 1.
 DR PROSITE: PS00404; SUBTILASE VAL; 1.
 DR PROSITE: PS00405; SUBTILASE THR; 1.
 DR PROSITE: PS00406; SUBTILASE LEU; 1.
 DR PROSITE: PS00407; SUBTILASE ILE; 1.
 DR PROSITE: PS00408; SUBTILASE MET; 1.
 DR PROSITE: PS00409; SUBTILASE PHE; 1.
 DR PROSITE: PS00410; SUBTILASE TYR; 1.
 DR PROSITE: PS00411; SUBTILASE ASP; 1.
 DR PROSITE: PS00412; SUBTILASE HIS; 1.
 DR PROSITE: PS00413; SUBTILASE SER; 1.
 DR PROSITE: PS00414; SUBTILASE TRP; 1.
 DR PROSITE: PS00415; SUBTILASE VAL; 1.
 DR PROSITE: PS00416; SUBTILASE THR; 1.
 DR PROSITE: PS00417; SUBTILASE LEU; 1.
 DR PROSITE: PS00418; SUBTILASE ILE; 1.
 DR PROSITE: PS00419; SUBTILASE MET; 1.
 DR PROSITE: PS00420; SUBTILASE PHE; 1.
 DR PROSITE: PS00421; SUBTILASE TYR; 1.
 DR PROSITE: PS00422; SUBTILASE ASP; 1.
 DR PROSITE: PS00423; SUBTILASE HIS; 1.
 DR PROSITE: PS00424; SUBTILASE SER; 1.
 DR PROSITE: PS00425; SUBTILASE TRP; 1.
 DR PROSITE: PS00426; SUBTILASE VAL; 1.
 DR PROSITE: PS00427; SUBTILASE THR; 1.
 DR PROSITE: PS00428; SUBTILASE LEU; 1.
 DR PROSITE: PS00429; SUBTILASE ILE; 1.
 DR PROSITE: PS00430; SUBTILASE MET; 1.
 DR PROSITE: PS00431; SUBTILASE PHE; 1.
 DR PROSITE: PS00432; SUBTILASE TYR; 1.
 DR PROSITE: PS00433; SUBTILASE ASP; 1.
 DR PROSITE: PS00434; SUBTILASE HIS; 1.
 DR PROSITE: PS00435; SUBTILASE SER; 1.
 DR PROSITE: PS00436; SUBTILASE TRP; 1.
 DR PROSITE: PS00437; SUBTILASE VAL; 1.
 DR PROSITE: PS00438; SUBTILASE THR; 1.
 DR PROSITE: PS00439; SUBTILASE LEU; 1.
 DR PROSITE: PS00440; SUBTILASE ILE; 1.
 DR PROSITE: PS00441; SUBTILASE MET; 1.
 DR PROSITE: PS00442; SUBTILASE PHE; 1.
 DR PROSITE: PS00443; SUBTILASE TYR; 1.
 DR PROSITE: PS00444; SUBTILASE ASP; 1.
 DR PROSITE: PS00445; SUBTILASE HIS; 1.
 DR PROSITE: PS00446; SUBTILASE SER; 1.
 DR PROSITE: PS00447; SUBTILASE TRP; 1.
 DR PROSITE: PS00448; SUBTILASE VAL; 1.
 DR PROSITE: PS00449; SUBTILASE THR; 1.
 DR PROSITE: PS00450; SUBTILASE LEU; 1.
 DR PROSITE: PS00451; SUBTILASE ILE; 1.
 DR PROSITE: PS00452; SUBTILASE MET; 1.
 DR PROSITE: PS00453; SUBTILASE PHE; 1.
 DR PROSITE: PS00454; SUBTILASE TYR; 1.
 DR PROSITE: PS00455; SUBTILASE ASP; 1.
 DR PROSITE: PS00456; SUBTILASE HIS; 1.
 DR PROSITE: PS00457; SUBTILASE SER; 1.
 DR PROSITE: PS00458; SUBTILASE TRP; 1.
 DR PROSITE: PS00459; SUBTILASE VAL; 1.
 DR PROSITE: PS00460; SUBTILASE THR; 1.
 DR PROSITE: PS00461; SUBTILASE LEU; 1.
 DR PROSITE: PS00462; SUBTILASE ILE; 1.
 DR PROSITE: PS00463; SUBTILASE MET; 1.
 DR PROSITE: PS00464; SUBTILASE PHE; 1.
 DR PROSITE: PS00465; SUBTILASE TYR; 1.
 DR PROSITE: PS00466; SUBTILASE ASP; 1.
 DR PROSITE: PS00467; SUBTILASE HIS; 1.
 DR PROSITE: PS00468; SUBTILASE SER; 1.
 DR PROSITE: PS00469; SUBTILASE TRP; 1.
 DR PROSITE: PS00470; SUBTILASE VAL; 1.
 DR PROSITE: PS00471; SUBTILASE THR; 1.
 DR PROSITE: PS00472; SUBTILASE LEU; 1.
 DR PROSITE: PS00473; SUBTILASE ILE; 1.
 DR PROSITE: PS00474; SUBTILASE MET; 1.
 DR PROSITE: PS00475; SUBTILASE PHE; 1.
 DR PROSITE: PS00476; SUBTILASE TYR; 1.
 DR PROSITE: PS00477; SUBTILASE ASP; 1.
 DR PROSITE: PS00478; SUBTILASE HIS; 1.
 DR PROSITE: PS00479; SUBTILASE SER; 1.
 DR PROSITE: PS00480; SUBTILASE TRP; 1.
 DR PROSITE: PS00481; SUBTILASE VAL; 1.
 DR PROSITE: PS00482; SUBTILASE THR; 1.
 DR PROSITE: PS00483; SUBTILASE LEU; 1.
 DR PROSITE: PS00484; SUBTILASE ILE; 1.
 DR PROSITE: PS00485; SUBTILASE MET; 1.
 DR PROSITE: PS00486; SUBTILASE PHE; 1.
 DR PROSITE: PS00487; SUBTILASE TYR; 1.
 DR PROSITE: PS00488; SUBTILASE ASP; 1.
 DR PROSITE: PS00489; SUBTILASE HIS; 1.
 DR PROSITE: PS00490; SUBTILASE SER; 1.
 DR PROSITE: PS00491; SUBTILASE TRP; 1.
 DR PROSITE: PS00492; SUBTILASE VAL; 1.
 DR PROSITE: PS00493; SUBTILASE THR; 1.
 DR PROSITE: PS00494; SUBTILASE LEU; 1.
 DR PROSITE: PS00495; SUBTILASE ILE; 1.
 DR PROSITE: PS00496; SUBTILASE MET; 1.
 DR PROSITE: PS00497; SUBTILASE PHE; 1.
 DR PROSITE: PS00498; SUBTILASE TYR; 1.
 DR PROSITE: PS00499; SUBTILASE ASP; 1.
 DR PROSITE: PS00500; SUBTILASE HIS; 1.
 DR PROSITE: PS00501; SUBTILASE SER; 1.
 DR PROSITE: PS00502; SUBTILASE TRP; 1.
 DR PROSITE: PS00503; SUBTILASE VAL; 1.
 DR PROSITE: PS00504; SUBTILASE THR; 1.
 DR PROSITE: PS00505; SUBTILASE LEU; 1.
 DR PROSITE: PS00506; SUBTILASE ILE; 1.
 DR PROSITE: PS00507; SUBTILASE MET; 1.
 DR PROSITE: PS00508; SUBTILASE PHE; 1.
 DR PROSITE: PS00509; SUBTILASE TYR; 1.
 DR PROSITE: PS00510; SUBTILASE ASP; 1.
 DR PROSITE: PS00511; SUBTILASE HIS; 1.
 DR PROSITE: PS00512; SUBTILASE SER; 1.
 DR PROSITE: PS00513; SUBTILASE TRP; 1.
 DR PROSITE: PS00514; SUBTILASE VAL; 1.
 DR PROSITE: PS00515; SUBTILASE THR; 1.
 DR PROSITE: PS00516; SUBTILASE LEU; 1.
 DR PROSITE: PS00517; SUBTILASE ILE; 1.
 DR PROSITE: PS00518; SUBTILASE MET; 1.
 DR PROSITE: PS00519; SUBTILASE PHE; 1.
 DR PROSITE: PS00520; SUBTILASE TYR; 1.
 DR PROSITE: PS00521; SUBTILASE ASP; 1.
 DR PROSITE: PS00522; SUBTILASE HIS; 1.
 DR PROSITE: PS00523; SUBTILASE SER; 1.
 DR PROSITE: PS00524; SUBTILASE TRP; 1.
 DR PROSITE: PS00525; SUBTILASE VAL; 1.
 DR PROSITE: PS00526; SUBTILASE THR; 1.
 DR PROSITE: PS00527; SUBTILASE LEU; 1.
 DR PROSITE: PS00528; SUBTILASE ILE; 1.
 DR PROSITE: PS00529; SUBTILASE MET; 1.
 DR PROSITE: PS00530; SUBTILASE PHE; 1.
 DR PROSITE: PS00531; SUBTILASE TYR; 1.
 DR PROSITE: PS00532; SUBTILASE ASP; 1.
 DR PROSITE: PS00533; SUBTILASE HIS; 1.
 DR PROSITE: PS00534; SUBTILASE SER; 1.
 DR PROSITE: PS00535; SUBTILASE TRP; 1.
 DR PROSITE: PS00536; SUBTILASE VAL; 1.
 DR PROSITE: PS00537; SUBTILASE THR; 1.
 DR PROSITE: PS00538; SUBTILASE LEU; 1.
 DR PROSITE: PS00539; SUBTILASE ILE; 1.
 DR PROSITE: PS00540; SUBTILASE MET; 1.
 DR PROSITE: PS00541; SUBTILASE PHE; 1.
 DR PROSITE: PS00542; SUBTILASE TYR; 1.
 DR PROSITE: PS00543; SUBTILASE ASP; 1.
 DR PROSITE: PS00544; SUBTILASE HIS; 1.
 DR PROSITE: PS00545; SUBTILASE SER; 1.
 DR PROSITE: PS00546; SUBTILASE TRP; 1.
 DR PROSITE: PS00547; SUBTILASE VAL; 1.
 DR PROSITE: PS00548; SUBTILASE THR; 1.
 DR PROSITE: PS00549; SUBTILASE LEU; 1.
 DR PROSITE: PS00550; SUBTILASE ILE; 1.
 DR PROSITE: PS00551; SUBTILASE MET; 1.
 DR PROSITE: PS00552; SUBTILASE PHE; 1.
 DR PROSITE: PS00553; SUBTILASE TYR; 1.
 DR PROSITE: PS00554; SUBTILASE ASP; 1.
 DR PROSITE: PS00555; SUBTILASE HIS; 1.
 DR PROSITE: PS00556; SUBTILASE SER; 1.
 DR PROSITE: PS00557; SUBTILASE TRP; 1.
 DR PROSITE: PS00558; SUBTILASE VAL; 1.
 DR PROSITE: PS0055

Best Local Similarity 80.0%; Pred. No. 31;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 56 CDADC 60

RESULT 19

Q9QZV3 PRELIMINARY; PRT; 76 AA.

AC Q9QZV3; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE ADAMS protein (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
STRAIN=WISTR; TISSUE=OVARY;
MEDLINE=20191571; PubMed=10727282;
RA Espey L.L., Yoshiooka S., Russell D.L., Robker R.L., Fujii S.,
RA Richards J.S.;
RT "Ovarian expression of a disintegrin and metalloproteinase with
RT thrombospondin motifs during ovulation in the gonadotropin-primed
RT immature rat.";
RT Biol. Reprod. 62:1090-1095(2000).
DR EMBL; AF159096; AAD56631.1; -.
FT NON TER 1 1
SQ SEQUENCE 76 AA; 9109 MW; F5FB320D7EF93912 CRC64;

Query Match 90.9%; Score 30; DB 11; Length 76;
Best Local Similarity 80.0%; Pred. No. 31;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 55 CDSEC 59

RESULT 20

Q9CWL2 PRELIMINARY; PRT; 209 AA.

AC Q9CWL2; 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 2410019P08Rik protein.
GN 2410019P08Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gastlisch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzairelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Sessa T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL; AK010559; BAB27027.1; -.
DR MGD; MGI:1916993; 2410019P08Rik.
DR InterPro; IPR000822; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
SQ SEQUENCE 209 AA; 22134 MW; 70D5FE056EBDF3D CRC64;

Query Match 90.9%; Score 30; DB 11; Length 209;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 62 CDAC 66

RESULT 21

Q99GR3 PRELIMINARY; PRT; 235 AA.

AC Q99GR3; 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative late expression factor 1 (CUN045 putative lef-1 late
DE expression factor 1, similar to ACPNPV ORF14).
GN CUN045.
OS Culex nigripalpus baculovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
OX NCBI_TaxID=130556;
[1]
SEQUENCE FROM N.A.
RA Moser B.A., Becnel J.J., White S.E., Afonso C., Kutish G., Shanker S.,
RA Almira E.;
RT "Morphological and molecular evidence that Culex nigripalpus
RT baculovirus is an unusual member of the family Baculoviridae.";
RT J. Gen. Virol. 82:283-297(2001).
[2]
SEQUENCE FROM N.A.
RA STRAIN=FLORIDA1997;
RC MEDLINE=2148685; PubMed=11602755;
RA Afonso C.J., Tullman E.R., Lu Z., Balinsky C.A., Moser B.A.,
RA Becnel J.J., Rock D.L., Kutish G.F.;
RT "Genome Sequence of a Baculovirus Pathogenic for Culex nigripalpus.";
RT J. Virol. 75:1157-1165(2001).
[3]
SEQUENCE FROM N.A.
RA STRAIN=FLORIDA1997;
RC Afonso C.J., Tullman E.R., Lu Z., Balinsky C.A., Moser B.A.,
RA Becnel J.J., Rock D.L., Kutish G.F.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF274283; AAK13260.1; -.
DR EMBL; AF403738; AAK94123.1; -.
SQ SEQUENCE 235 AA; 26871 MW; 4B9CCD21D9478B27 CRC64;

Query Match 90.9%; Score 30; DB 12; Length 235;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAC 5
Db 126 CDADC 130

RESULT 22
Q21622 PRELIMINARY; PRT; 245 AA.

AC Q21622;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE R01H10.4 protein.
 GN R01H10.4
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lightning J., Lloyd C.R.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 DR "Genome sequence of the nematode C.elegans: A platform for
 investigating biology.";
 S Q EMBL; Z31590; CAA83462.1; -;
 S Q SEQUENCE 245 AA; 27219 MW; F26D9B616163463D CRC64;

Query Match 90.9%; Score 30; DB 5; Length 245;
 Best Local Similarity 80.0%; Pred. No. 80;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 205 CDAQC 209

RESULT 23

Q9PT11 PRELIMINARY; PRT; 341 AA.
 ID Q9PT11;
 AC Q9PT11;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Galectin like protein.
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OC NCBI_TaxID=8022;
 RN [1]
 GN SEQUENCE FROM N.A.
 TTSSUB=HEAD KIDNEY;
 RA Soma G., Inagawa H., Honda T., Nakanishi T., Otake M., Nishizawa T.,
 RA Kanou J., Endo M.;
 RT "Rainbow trout cDNA similar to galectin 9 like."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DE EMBL; AB027452; BAA88670.1; -;
 DR HSSP; P47929; 1BKZ.
 DR InterPro; IPR001079; Galectin.
 DR Pfam; PF00337; Gal_bind_lectin; 2.
 DR SMART; SM00276; GLECT; 2.
 DR PROSITE; PS00309; GALACTIN; UNKNOWN 1.
 S Q SEQUENCE 341 AA; 37816 MW; 176D535EC19550C4 CRC64;

Query Match 90.9%; Score 30; DB 13; Length 341;
 Best Local Similarity 80.0%; Pred. No. 1e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 298 CDAQC 302

RESULT 24

Q9FGH5 PRELIMINARY; PRT; 363 AA.
 Q9FGH5

AC Q9FGH5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Genomic DNA, chromosome 5, pl clone:MOJ2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DE EMBL; AB025632; BAB10262.1; -;
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR001876; Znf_RandDP.
 DR Pfam; PF00076; rrm; 1.
 DR Pfam; PF00641; zf-RanBP; 1.
 DR PRINTS; PR01228; EGGSHLL.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; Znf_RBZ; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS01358; ZF_RANBP2_1; UNKNOWN 1.
 S Q SEQUENCE 363 AA; 36150 MW; F9FCBBE3EE1D42FC CRC64;

Query Match 90.9%; Score 30; DB 10; Length 363;
 Best Local Similarity 80.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 269 CDAQC 273

RESULT 25

P91419 PRELIMINARY; PRT; 395 AA.
 ID P91419;
 AC P91419;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hypothetical 42.8 kDa protein.
 GN T01D1.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium."
 RL Science 282:2012-2018(1998).
 S Q SEQUENCE FROM N.A.
 RP STRAIN=BRISTOL N2;
 RC Bradsen H., Wohlmann P.;
 RT "The sequence of C. elegans cosmid T01D1."
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submision."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DE EMBL; U80455; AAB37887.2; -;

DR InterPro: IPR003341; DUF139.
 DR Pfam: PF02363; DUF139; 8.
 KW Hypothetical protein.
 SQ SEQUENCE 395 AA; 42801 MW; E7B7C7878773A6BA CRC64;

Query Match 90.9%; Score 30; DB 5; Length 395;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 DB 321 CDAQC 325

RESULT 26
 ID 013455 PRELIMINARY; PRT; 416 AA.
 AC 013455;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 RT beta-glucanase from *Aspergillus oryzae* KRN616.
 RL Appl. Microbiol. Biotechnol. 46:538-544 (1996).
 DR EMBL; D83732; BAA22589.1; -.
 DR HSSP; P07981; 1EG1.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD186135; GH_7; 1.
 KW Glycosidase; Hydrolase.
 SQ SEQUENCE 416 AA; 44462 MW; A1APC9B560CE0E19 CRC64;

OS *Aspergillus oryzae*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.
 NCBI_TaxID=5062;

RP SEQUENCE FROM N.A.
 RC STRAIN=KRN616;
 RX MEDLINE=97161783; PubMed=9008887;
 RA Kitamoto N., Go M., Shibayama T., Kimura T., Kito Y., Ohmura K.,
 RA Tsukagoshi N.;
 RT "Molecular cloning, purification and characterization of two endo-1,4-
 RT beta-glucanases from *Aspergillus oryzae* KRN616."
 RL Appl. Microbiol. Biotechnol. 46:538-544 (1996).

DR HSSP; P07981; 1EG1.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD186135; GH_7; 1.
 KW Glycosidase; Hydrolase.
 SQ SEQUENCE 416 AA; 44462 MW; A1APC9B560CE0E19 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 416;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 1 CDAEC 5
 189 CDAQC 193

RESULT 27
 ID 094KD0 PRELIMINARY; PRT; 422 AA.
 AC 094KD0;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 RT beta-glucanase from *Aspergillus oryzae* KRN616.
 RL Appl. Microbiol. Biotechnol. 46:538-544 (1996).
 DR EMBL; A7598470/mqj2.60.
 DR AT598470/mqj2.60.
 OS *Aspergillus thaliana* (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Shim P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bower L., Carrinci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
 RA Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,

RA Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
 RA Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
 RA Shim P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bower L., Carrinci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF367294; AA56282.1; -.
 DR EMBL; AY075677; AB177684.1; -.
 DR InterPro: IPR000504; RNA_rec_mct.
 DR InterPro: IPR01876; Znf_RangDP.
 DR Pfam: PF00076; rrm; 1.
 DR Pfam: PF00641; zf-RanBP; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS01358; ZF_RANBP2_1; UNKNOWN_1.
 DR PROSITE; PS01358; ZF_RANBP2_1; UNKNOWN_1.
 SQ SEQUENCE 422 AA; 42316 MW; 0ED75C7F73575133 CRC64;

Query Match 90.9%; Score 30; DB 10; Length 422;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 269 CDAQC 273

RESULT 28
 ID 09UVP4 PRELIMINARY; PRT; 423 AA.
 AC 09UVP4;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 RT beta-glucanase from *Alternaria alternata*.
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF176571; AAF05699.1; -.
 DR HSSP; P00725; 8CEL.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD186135; GH_7; 1.
 SQ SEQUENCE 423 AA; 45525 MW; B7E08851668F3068 CRC64;

OS *Alternaria alternata*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; *Alternaria*.
 NCBI_TaxID=5599;

RP SEQUENCE FROM N.A.
 RA Behl D., Prusky D., Dinooor A.;
 RT "Exoglucanase of *Alternaria alternata*."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF176571; AAF05699.1; -.
 DR HSSP; P00725; 8CEL.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom; PD186135; GH_7; 1.
 SQ SEQUENCE 423 AA; 45525 MW; B7E08851668F3068 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 423;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 161 CDAQC 165

RESULT 29
 ID 09C1F9 PRELIMINARY; PRT; 423 AA.
 AC 09C1F9;

DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Cellulase.
 GN CEL2.
 OS Cochliobolus carbonum (Bipolaris zeicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; Cochliobolus.
 OX NCBI_Taxid=5017;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21364349; PubMed=11471744;
 RA Ann J.-H., Sposato P., Kim S.I., Walton J.D.,
 RT "Molecular cloning and characterization of cel2 from the fungus
 RT Cochliobolus carbonum."
 RL Biosci. Biotechnol. Biochem. 65:1406-1411(2001).
 DR EMBL: AF36799; AKK19621.1; -.
 DR HSSP: P56680; 1A39.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR Prodom: PD186135; GH_7; 1.
 SO SEQUENCE 423 AA; 45824 MW; 17A0E0F91913D1DF CRC64;

Query Match 90.9%; Score 30; DB 3; Length 423;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 Db 188 CDAQC 192

RESULT 30
 Q9HGT3 PRELIMINARY; PRT; 429 AA.
 AC O9HGT3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Endoglucanase (EC 3.2.1.4).
 GN EG-1.
 OS Fusarium oxysporum f. sp. lycopersici.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Fusarium.
 OX NCBI_Taxid=59765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX Nishino O.T., Warren R.A.J.,
 RT "Cloning, sequencing and expression of the cellulase genes of Fusarium
 RT oxysporum f. sp. lycopersici."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF292710; AAG09047.1; -.
 DR HSSP: P46237; 30WV.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR Prodom: PD186135; GH_7; 1.
 DR GlycoSite: Hydrolase.
 KW SEQUENCE 429 AA; 46466 MW; AAE3B0776EB34791 CRC64;
 SO

Query Match 90.9%; Score 30; DB 3; Length 429;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 Db 190 CDAQC 194

RESULT 31
 Q9P8K8 PRELIMINARY; PRT; 438 AA.
 AC Q9P8K8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Cellulase CEL1.
 GN CEL1.
 OS Leptosphaeria maculans (Blackleg fungus).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Leptosphaeriaceae; Leptosphaeria.
 OX NCBI_Taxid=5022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20267851; PubMed=10806354;
 RA Sexton A.C., Paulsen M., Woestemeyer J., Howlett B.J.,
 RT "Cloning, characterization and chromosomal location of three genes
 RT encoding host cell wall-degrading enzymes in Leptosphaeria maculans, a
 RT fungal pathogen of Brassica spp."
 RL Gene 248:89-97(2000).
 DR EMBL: AF240000; AAF70170.1; -.
 DR HSSP: P00725; 6CEL.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR Prodom: PD186135; GH_7; 1.
 SO SEQUENCE 438 AA; 47623 MW; 8D855EB31F434960 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 438;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 Db 177 CDAQC 181

RESULT 32
 Q9P8K7 PRELIMINARY; PRT; 447 AA.
 AC Q9P8K7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Cellulase CEL2.
 GN CEL2.
 OS Leptosphaeria maculans (Blackleg fungus).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Leptosphaeriaceae; Leptosphaeria.
 OX NCBI_Taxid=5022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20267851; PubMed=10806354;
 RA Sexton A.C., Paulsen M., Woestemeyer J., Howlett B.J.,
 RT "Cloning, characterization and chromosomal location of three genes
 RT encoding host cell wall-degrading enzymes in Leptosphaeria maculans, a
 RT fungal pathogen of Brassica spp."
 RL Gene 248:89-97(2000).
 DR EMBL: AF240001; AAF70171.1; -.
 DR HSSP: P00725; 6CEL.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR Prodom: PD186135; GH_7; 1.
 SO SEQUENCE 447 AA; 48552 MW; 970A582ED18676EA CRC64;

Query Match 90.9%; Score 30; DB 3; Length 447;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 Db 188 CDAQC 192

RESULT 33
 Q95YH1 PRELIMINARY; PRT; 450 AA.
 AC Q95YH1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Cellulase.
 GN PGCBH-HOMO1.
 OS Pseudotriconympha grisea.
 OC Eukaryota; Parabasalida; Hypermastigida; Trichonymphidae;
 OC Pseudotriconympha.
 OX NCBI_TaxID=104083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakashima K., Watanabe H., Azuma J.;
 RT "Cellulase genes from the parabasalid symbiont Pseudotriconympha
 grisea in the hindgut of a wood-feeding termite, Coptotermes
 formosanus";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB071864; BAB69425.1; -.
 DR InterPro: IPR001722; GH 7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom: PD186135; GH 7; 1.
 DR SEQUENCE 450 AA; 48934 MW; D67625AD6C3BB9 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 450;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 184 CDAQC 188

RESULT 34
 O95NR5 PRELIMINARY; PRT; 450 AA.
 AC O95NR5;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Cellulase.
 GN PGCBH-HOMO2 OR PGCBH-HOMO2A OR PGCBH-HOMO2B OR PGCBH-HOMO2C.
 OS Pseudotriconympha grisea.
 OC Eukaryota; Parabasalida; Hypermastigida; Trichonymphidae;
 OC Pseudotriconympha.
 OX NCBI_TaxID=104083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakashima K., Watanabe H., Azuma J.;
 RT "Cellulase genes from the parabasalid symbiont Pseudotriconympha
 grisea in the hindgut of a wood-feeding termite, Coptotermes
 formosanus";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB071868; BAB69429.1; -.
 DR EMBL: AB071865; BAB69426.1; -.
 DR EMBL: AB071866; BAB69427.1; -.
 DR EMBL: AB071867; BAB69428.1; -.
 DR InterPro: IPR001722; GH 7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom: PD186135; GH 7; 1.
 DR SEQUENCE 450 AA; 48771 MW; 718B1ED6A60DEBA CRC64;

Query Match 90.9%; Score 30; DB 5; Length 450;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 184 CDAQC 188

RESULT 35
 O9UUS1 PRELIMINARY; PRT; 451 AA.
 ID O9UUS1;
 AC O9UUS1;

DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Cellulase 1,4-beta-cellobiosidase (EC 3.2.1.91) (Fragment).
 GN CBH1.2.
 OS Humicola grisea var. thermoides.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=5528;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pocas-Fonseca M.J., Rocha B.B., Silva-Pereira I., Azevedo M.O.;
 RT "RT-PCR obtained cbh1.2 cDNA clone from Humicola grisea var.
 thermoides";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF123441; AAD31545.1; -.
 DR HSSP: P00725; 8CEL.
 DR InterPro: IPR001722; GH 7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom: PD186135; GH 7; 1.
 DR Glycosidase; Hydrolase.
 DR NON TER 451
 DR SEQUENCE 451 AA; 49654 MW; F5BF0F054E73B3B1 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 451;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 192 CDAQC 196

RESULT 36
 O94093 PRELIMINARY; PRT; 451 AA.
 ID O94093;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE 1,4-beta-D-glucan cellobiohydrolase (EC 3.2.1.91).
 GN CBH1.2.
 OS Humicola grisea.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=5527;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THERMOIDEA;
 RA Pocas-Fonseca M.J., Lima B.D., Brigido M.M., Pereira I.S.,
 RA Felipe M.S.S., Radford A., Azevedo M.O.;
 RT "Humicola grisea var. thermoides cbh1.2: A new gene in the family of
 cellobiohydrolases is expressed and encodes a cellulose-binding
 domain-less enzyme";
 RL J. Gen. Appl. Microbiol. 43:115-120 (1997).
 DR EMBL: U50594; AAD11942.1; -.
 DR HSSP: P00725; 8CEL.
 DR InterPro: IPR001722; GH 7.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR ProDom: PD186135; GH 7; 1.
 DR Glycosidase; Hydrolase.
 DR SEQUENCE 451 AA; 49654 MW; F5BF0F054E73B3B1 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 451;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
 DB 192 CDAQC 196

RESULT 37
 O93780 PRELIMINARY; PRT; 451 AA.
 ID O93780

```

AC 093780;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Exoglucanase.
GN EXO1.
OS Humicola grisea var. thermoides.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
ON NCBI_TaxID=5528;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO9854;
RX MEDLINE=98429588; PubMed=9756616;
RT Takashima S., Iikura H., Nakamura A., Hidak M., Masaki H., Uozumi T.;
RT "Isolation of the gene and characterization of the enzymatic
RT properties of a major exoglucanase of Humicola grisea without a
RT cellulose-binding domain."
RL U. Biochem. 124:717-725(1998).
DR EMBL; AB003105; BAA74517.1; -.
DR HSSP; P00725; BCEL.
DR InterPro: IPR001722; GH_7.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD186135; GH_7.1.
SQ SEQUENCE 451 AA; 49655 MW; F5BF0F082313B3B1 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 451;
Best Local Similarity 80.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 192 CDAQC 196

RESULT 38
Q12621 PRELIMINARY; PRT; 525 AA.
ID Q12621
AC Q12621;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulase (EC 3.2.1.91).
GN CBH-1.
OS Humicola grisea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
ON NCBI_TaxID=5527;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO9854;
RT "Cloning, sequencing, and expression of the cellulase genes of
RT Humicola grisea var. thermoides."
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; D63515; BAA09785.1; -.
DR HSSP; P00725; IAZ6.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001722; GH_7.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; GH_fungal; 1.
DR ProDom: PD186135; GH_7.1.
DR SMART; SM00236; fCB1.1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 525 AA; 55721 MW; A2E6SF40F6D3B50 CRC64;

Query Match 90.9%; Score 30; DB 3; Length 525;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 191 CDAQC 195

```

```

RESULT 39
Q96134 PRELIMINARY; PRT; 528 AA.
ID Q96134
AC Q96134;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Unknown (Protein for MGC:14333) (Hypothetical 57.8 kDa protein).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007854; AAH07854.1; -.
DR EMBL; BC015896; AAH15896.1; -.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001230; Prenyl_site.
DR InterPro: IPR000408; Reg_chf_condens.
DR Pfam: PF00023; ank; 5.
DR PROSITE; PS50068; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
KW Hypothetical protein; ANK repeat; Repeat.
SQ SEQUENCE 528 AA; 57811 MW; A5A2D5E1A16A438C CRC64;

Query Match 90.9%; Score 30; DB 4; Length 528;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDAEC 5
Db 133 CDESC 137

RESULT 40
Q9V821 PRELIMINARY; PRT; 613 AA.
ID Q9V821
AC Q9V821;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG13872 protein.
GN CG13872.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phyllophaga; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```


RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacled J.M.,
 RA Palazolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtikas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 DR EMBL; AE003794; AAE5513.1;-
 DR FlyBase; FBgn0034477; CG13872.
 DR SEQUENCE 613 AA; 69069 MM; 69DA9BC1F5575D8C CRC64;

Query Match 90.9%; Score 30; DB 5; Length 613;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CDAEC 5
 Db 365 CDSEC 369

Search completed: July 11, 2003, 15:31:16
 Job time : 81 secs

THIS PAGE BLANK (USPTO)